

CD4⁺ T cells exhibit distinct transcriptional phenotypes in the lymph nodes and blood following mRNA vaccination in humans

Received: 28 August 2023

Accepted: 6 June 2024

Published online: 20 August 2024

 Check for updates

Nicholas Borcherding¹, Wooseob Kim^{1,2}, Michael Quinn³, Fangjie Han⁴, Julian Q. Zhou¹, Alexandria J. Sturtz¹, Aaron J. Schmitz¹, Tingting Lei¹, Stefan A. Schattgen¹, Michael K. Klebert⁶, Teresa Suessen⁷, William D. Middleton⁷, Charles W. Goss⁸, Chang Liu¹, Jeremy Chase Crawford¹, Paul G. Thomas¹, Sharlene A. Teeffey⁷, Rachel M. Presti^{6,9,10,11}, Jane A. O'Halloran⁹, Jackson S. Turner¹, Ali H. Ellebedy^{1,10,11}✉ & Philip A. Mudd¹,[✉]

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and mRNA vaccination induce robust CD4⁺ T cell responses. Using single-cell transcriptomics, here, we evaluated CD4⁺ T cells specific for the SARS-CoV-2 spike protein in the blood and draining lymph nodes (dLNs) of individuals 3 months and 6 months after vaccination with the BNT162b2 mRNA vaccine. We analyzed 1,277 spike-specific CD4⁺ T cells, including 238 defined using Trex, a deep learning-based reverse epitope mapping method to predict antigen specificity. Human dLN spike-specific CD4⁺ follicular helper T (T_{FH}) cells exhibited heterogeneous phenotypes, including germinal center CD4⁺ T_{FH} cells and CD4⁺IL-10⁺ T_{FH} cells. Analysis of an independent cohort of SARS-CoV-2-infected individuals 3 months and 6 months after infection found spike-specific CD4⁺ T cell profiles in blood that were distinct from those detected in blood 3 months and 6 months after BNT162b2 vaccination. Our findings provide an atlas of human spike-specific CD4⁺ T cell transcriptional phenotypes in the dLNs and blood following SARS-CoV-2 vaccination or infection.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic provided a unique opportunity to study primary human immune responses to a new pathogen and the immunodominant SARS-CoV-2 spike antigen incorporated into various vaccine platforms. mRNA vaccines engender strong immune responses to the SARS-CoV-2 spike antigen, including high-frequency circulating spike-specific CD4⁺ T cells^{1,2} and spike-specific CD4⁺ follicular helper T (T_{FH}) cells in the draining lymph nodes (dLNs)³. CD4⁺ T_{FH} cells support the development and maintenance of germinal center (GC) B cells in secondary lymphoid organs by providing appropriate co-stimulation and cytokine survival signals throughout antibody class switch, affinity maturation,

long-lived plasma cell development and memory B cell development^{4–6}. In mouse models, functional CD4⁺ T_{FH} cells are absolutely required for productive GC responses and the development of memory B cells and long-lived plasma cells^{7–11}.

Due to limitations in sampling human secondary lymphoid organs, GC and CD4⁺ T_{FH} cell responses have been studied in easily accessible tissue compartments, including blood and discarded clinical tonsillectomy tissue^{12–14}. Biopsies or autopsy samples have yielded insights into the phenotype of human CD4⁺ T_{FH} cells in the LNs^{15,16} but have been limited to the exploration of phenotypes at steady state. The evaluation of human antigen-specific CD4⁺ T_{FH} cells in the secondary

A full list of affiliations appears at the end of the paper. ✉ e-mail: ellebedy@wustl.edu; pmudd@wustl.edu

lymphoid organs after acute infection or vaccination is even more limited^{3,13,17} and rarely includes analysis of antigen-specific responses at the single-cell level.

Serial fine needle aspiration (FNA) of ultrasound-localized dLN_s has been used to probe human GC responses in the axillary dLN_s after deltoid intramuscular vaccination in a cohort of infection-naïve individuals vaccinated with SARS-CoV-2 mRNA vaccines^{3,18} and detected strong induction of spike-specific CD4⁺T_{FH} cell responses in the dLN_s, including CD4⁺T cells that recognize the HLA-DPB1*04-restricted immunodominant epitope S₁₆₇₋₁₈₀ (ref. 3) and persist for 6 months after vaccination³. Here, we performed single-cell RNA-sequencing (RNA-seq) to obtain matched transcriptome and T cell antigen receptor (TCR) sequencing from >200,000 T cells from the blood and dLN_s of six SARS-CoV-2-naïve, HLA-DPB1*04⁺ individuals on days 28–201 after vaccination with the first dose of a primary two-dose BNT162b2 mRNA vaccine series, with the second dose delivered exactly 21 days after the first dose. Using a reverse epitope discovery technique developed to integrate biochemical properties of TCR complementarity determining region 3 (CDR3) amino acids and transcriptional profiles in single cells to predict antigen specificity, we expanded the number of known spike-specific TCRs in our dataset, confirmed that these paired TCRs were spike-specific and analyzed the transcriptional dynamics of multiple lineages of spike-specific CD4⁺T cells restricted by multiple class II HLA alleles in the blood and dLN_s at days 110 and 201 after vaccination. We also incorporated the analysis of spike-specific CD4⁺T cells from the blood of a cohort of six HLA-DPB1*04⁺ individuals after primary infection with SARS-CoV-2 and compared these responses to the spike-specific memory CD4⁺T cells found after vaccination. Our data provide an atlas of total and spike-specific transcriptional phenotypes in CD4⁺T cells found in blood and the dLN_s following initial exposure to the SARS-CoV-2 spike antigen.

Results

mRNA vaccination induces diverse T cell phenotypes

We performed single-cell RNA-seq and paired TCR sequencing on total dLN cells from FNA samples obtained on days 28, 60, 110 and 201 after the first BNT162b2 vaccine dose from six 34- to 48-year-old female ($n = 2$) and male ($n = 4$) HLA-DPB1*04⁺ individuals and magnetically enriched total CD4⁺ cells from temporally matched blood samples from four of the six individuals obtained at days 110 and 201 after the first dose of the two-dose BNT162b2 vaccination^{3,19} (Fig. 1a and Extended Data Tables 1 and 2). At all time points, all participants had detectable spike-specific GC B cells in the evaluated dLN¹⁹. A total of 219,283 individual T cells passed all transcriptional quality metrics and contained a paired TCR sequence (Fig. 1b; <https://cellpilot.emed.wustl.edu>). Because we did not select for CD4⁺T cells during dLN sample preparation and because of a low level of contamination with CD8⁺T cells during the magnetic separation step in the blood samples (Extended Data Fig. 1), CD8⁺T cells were included in the dataset. Based on uniform manifold approximation and projection (UMAP) analysis, we identified 19 transcriptional T cell clusters (Fig. 1b). Following annotation using granular cell types with canonical markers and reference atlases (Extended Data Fig. 2), we identified two CD4⁺T_{FH} cell clusters (C10 and C15) and one CD4⁺follicular memory T (T_{FM}) cell cluster (C1) that colocalized in the same region of the UMAP (Fig. 1b). Common T cell markers, such as CD4, CD8A, CXCR5, ICOS and PDCD1, separated clearly in the UMAP projection (Fig. 1c). All 19 T cell clusters were present at all time points (Fig. 1d) and in both the blood and dLN_s (Fig. 1e). Throughout the UMAP projection, we found individual T cells with published SARS-CoV-2-specific TCR CDR3 sequences^{3,20,21}, including the immunodominant HLA-DPB1*04-restricted CD4⁺T cell epitope S₁₆₇₋₁₈₀ (ref. 3; Fig. 1f). S₁₆₇₋₁₈₀-specific TCRs were primarily localized in the two CD4⁺T_{FH} cell clusters (C10 and C15; Fig. 1b,f), consistent with an enrichment of dLN tissue from HLA-DPB1*04⁺ individuals during an ongoing GC response in our dataset. Alignment of previously

published spike-specific TCR α-chain (TRA) and TCR β-chain (TRB) CDR3 sequences^{3,20,21} found in our dataset identified dominant polar amino acid signatures at positions 11, 12 and 13 of the TRA CDR3 and positions 10 and 13 of the TRB CDR3 and similarities in amino acids found at other key CDR3 residues (Fig. 1g). As such, our analysis of more than 200,000 T cells found in the blood and dLN_s in the first 6 months following primary BNT162b2 mRNA vaccination revealed diverse T cell transcriptional profiles enriched for spike-specific CD4⁺T_{FH} cells.

dLN_s contain transcriptionally diverse CD4⁺T_{FH} cells

To analyze the phenotypic dynamics of human CD4⁺T_{FH} cells in the dLN_s after vaccination, we generated a new UMAP of all dLN CD4⁺T_{FH} cells and CD4⁺T_{FM} cells found in clusters C1, C10 and C15, which identified 12 distinct phenotypic subclusters (denoted c0–c11; Fig. 2a–c; <https://cellpilot.emed.wustl.edu>). These were principally classified as the well-described^{4,22} CXCL13⁺CXCR5⁺BCL6⁺CD4⁺GC T_{FH} cells (c3), the previously described^{15,23} CD4⁺interleukin-10⁺(CD4⁺IL-10⁺)T_{FH} cells (c8), cytotoxic GZMA⁺GZMK⁺CD4⁺T_{FH} (T_{FH}_C) cells (c9), effector IRF4⁺CD4⁺T_{FH} (T_{FH}_{EFF}) cells (c6), proliferating MKI67⁺CD4⁺T_{FH} (T_{FH}_{PRO}) cells (c11), follicular regulatory FOXP3⁺CD4⁺T cells (c4) and memory KLF2⁺CD4⁺T_{FM} cells (c0, c1, c2, c7 and c10). A single distinct cluster (c5) represented CD8⁺T cells and was not considered further. CD4⁺GC T_{FH} and CD4⁺IL-10⁺T_{FH} cells clustered together in hierarchical clustering analysis of gene sets (Fig. 2d) and shared many characteristics, including the highest expression of canonical CD4⁺GC T_{FH} markers like CXCR5, PDCD1 and BCL6 (Fig. 2b,c); the highest expression of genes related to TCR signaling, helper T cell pathways, activation pathways, cell adhesion signaling and antigen presentation in gene set enrichment analysis (Fig. 2d) and high expression of genes related to increased metabolic activity, with elevated expression levels of genes involved in oxidative phosphorylation, glycolysis and PI3K-AKT signaling (Fig. 2d). CD4⁺GC T_{FH} and CD4⁺IL-10⁺T_{FH} cells maintained a relatively consistent expression of the distinguishing gene sets throughout the duration of the GC response from days 28 to 201 (Fig. 2e) and shared the largest number of identical paired TCR clonotypes among all 12 dLN T_{FH} cell clusters (Fig. 2f), suggesting substantial overlap in the clonal populations recruited to these effector CD4⁺T_{FH} cell subsets. Despite the close relationship between CD4⁺GC T_{FH} and CD4⁺IL-10⁺T_{FH} cells, they exhibited differences in cytokine gene expression, with exclusive expression of IL10 and much higher expression of IL21 in CD4⁺IL-10⁺T_{FH} cells (c8) and much higher expression of IL4 in CD4⁺GC T_{FH} cells (c3; Fig. 2b).

Both CD4⁺T_{FH}_{EFF} cells and CD4⁺T_{FH}_C cells expressed much less CXCR5, PDCD1 and BCL6 than CD4⁺GC T_{FH} cells and CD4⁺IL-10⁺T_{FH} cells but segregated with CD4⁺GC T_{FH} and CD4⁺IL-10⁺T_{FH} cells in hierarchical clustering analysis (Fig. 2d), although the functional importance of these two T_{FH} populations was not clear. CD4⁺T_{FH}_{EFF} cells showed high expression of the microRNA MIR155HG (Fig. 2c), a transcript associated with increased inflammation through inhibition of SOCS1 and many other genes²⁴ that also encodes the short functional peptide miPEP155 that modulates class II antigen presentation²⁵, had the highest expression of the T cell transcriptional regulators IRF4 and NFKB1D (Fig. 2c), suggesting a transitional phenotype and that they may ultimately develop into other CD4⁺T cell subsets, and had the highest frequency of clonotypic overlap with CD4⁺GC T_{FH} cells (Fig. 2f), suggesting a relationship between these two subsets. CD4⁺T_{FH}_C cells exhibited clonal overlap with CD4⁺T_{FM} cells (c1) at days 28, 60, 110 and 201 and had minimal clonal overlap with the other clusters (Fig. 2f), suggesting a unique lineage distinct from CD4⁺GC T_{FH} cells that may be related to previously identified ‘cytotoxic’ T_{FH} cells¹³.

CD4⁺T_{FH}_{PRO} cells (c11) expressed the proliferation marker MKI67 (Fig. 2c) and expressed PDCD1 and CXCR5 (Fig. 2c). Although a spatiotemporal relationship between CD4⁺T_{FH}_{PRO} cells and the other effector T_{FH} cells could not be established from these data alone, unique paired TCR clonotypes were shared specifically between CD4⁺T_{FH}_{PRO}, CD4⁺GC T_{FH} and CD4⁺IL-10⁺T_{FH} cells at days 28, 60, 110 and

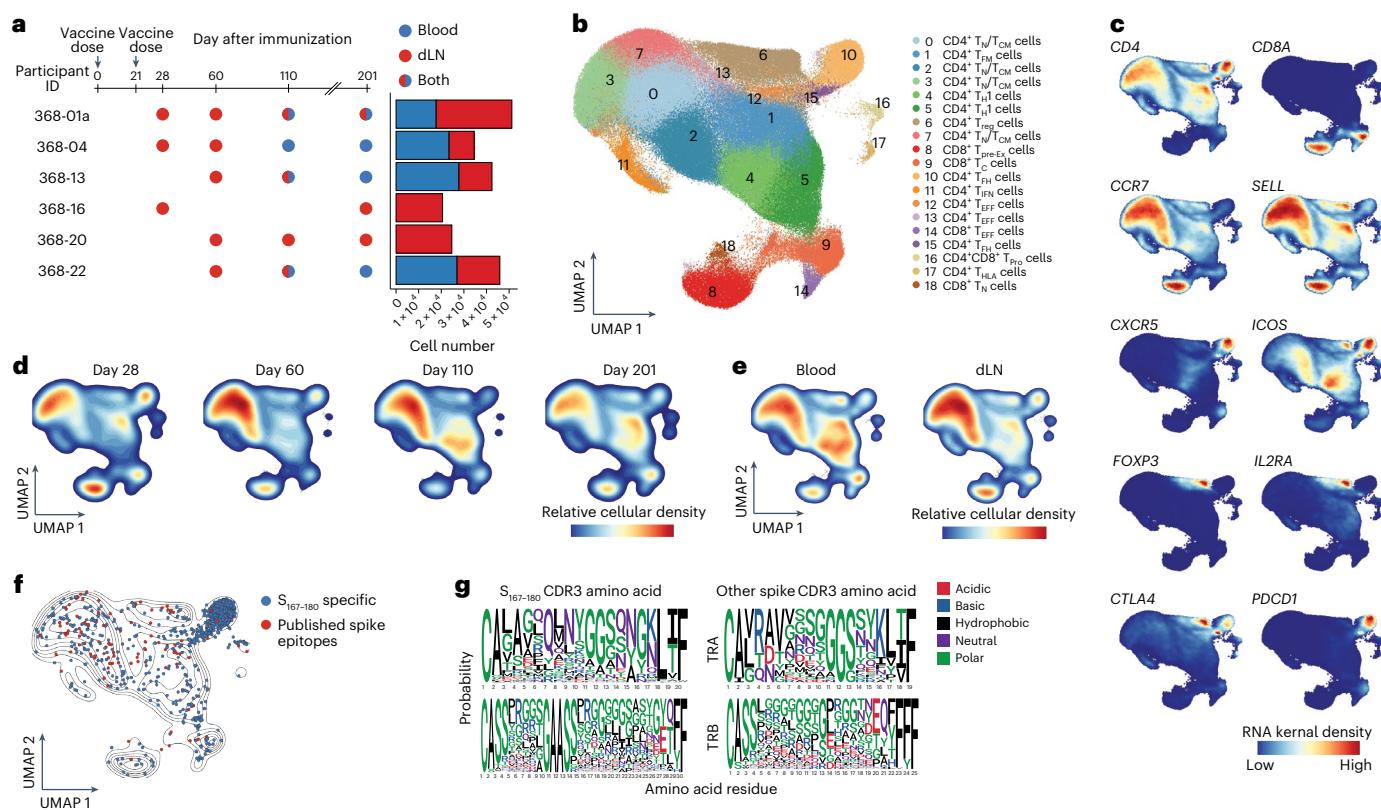


Fig. 1 | BNT162b2 mRNA vaccination induces spike-specific CD4⁺ T cells with diverse transcriptional phenotypes in the blood and dLNs. a, Schematic showing sample collection time points for dLN FNA and peripheral blood collection from six donors (368-01a, 368-04, 368-13, 368-16, 368-20 and 368-22) at days 28, 60, 110 and 201 after the first dose of the BNT162b2 vaccine (left) and graph showing the number of cells isolated from each donor from the dLN or blood (right). For each sample collection, a technical replicate was performed and sequenced. **b**, UMAP of 219,283 dLN and blood CD3⁺ T cells with paired TRA–TRB sequences that passed quality control filtering. Cluster annotation was based on canonical subtype markers and automated annotation using SingleR

and ProjectTIL; T_N, naive T cells; T_{H1}, type 1 helper T cells; T_{pre-Ex}, pre-exhausted T cells; T_C, cytotoxic T cells; T_{IFN}, IFN-expressing T cells; T_{EFF}, effector T cells; T_{Pro}, proliferating T cells; T_{HLA}, high HLA/MHC expression T cells; T_{CM}, central memory T cells. **c**, Gene-weighted density estimates overlaid on the UMAP coordinates for the T cell markers CD4, CD8A, CCR7, SELL, CXCR5, ICOS, FOXP3, IL2RA, CTLA4 and PDCD1. **d,e**, Relative cellular density at days 28, 60, 110 and 201 (**d**) and in the blood and dLN (**e**) in CD3⁺ T cells, as in **b**. **f**, Localization of spike-specific TCRs identified in refs. 3,21 along the UMAP. S₁₆₇₋₁₈₀-specific TCRs are highlighted in blue, and other spike-specific TCRs are in red. **g**, Alignment of TRA and TRB CDR3 motifs for S₁₆₇₋₁₈₀-specific³ and other spike-specific²¹ TCRs.

201 (Fig. 2f), suggesting that CD4⁺ GC T_{IFN} and CD4⁺ IL-10⁺ T_{IFN} clones entered proliferation cycles throughout the 6-month timeline rather than a burst of proliferation early and maintenance of the cluster size over time.

dLN CD4⁺ T_{FM} cells (c0, c1, c2, c7 and c10) had relatively low expression of genes involved in oxidative and glycolytic metabolism pathways and TCR signaling and cell adhesion signaling compared to the effector T_{IFN} populations (Fig. 2d) and expressed transcription factors involved in maintaining long-term T cell responsiveness and homeostasis, such as KLF2 (refs. 26,27), JUN²⁸, JUNB²⁹ and KLF6 (refs. 27; Fig. 2c). Cluster c7 and cluster c1 CD4⁺ T_{FM} cells exhibited clonal overlap with CD4⁺ GC T_{IFN} cells (Fig. 2f), suggesting a close relationship between CD4⁺ GC T_{IFN} cells and the CD4⁺ T_{FM} populations found in c7 and c1. TCR sequences specific for the S₁₆₇₋₁₈₀ epitope³ were found primarily in the CD4⁺ GC T_{IFN} cell cluster at every evaluated time point (Fig. 2g) and were found in the CD4⁺ IL-10⁺ T_{IFN}, CD4⁺ T_{IFN} and CD4⁺ T_{FM} cell clusters in small numbers at at least one time point (Fig. 2g). Thus, CD4⁺ T_{IFN} cell populations found in the dLN following BNT162b2 mRNA vaccination were transcriptionally diverse and included effector, memory, proliferating and regulatory populations and spike-specific CD4⁺ GC T_{IFN} cells.

Trex can identify antigen-specific CD4⁺ T_{IFN} cells

To expand the number of spike-specific TCR sequences in the present dataset, we developed a method (referred to as Trex (T cell receptor

and expression)) that used coembedding of the RNA transcriptome and the latent dimensional embeddings of both the TRA and TRB CDR3 sequences for each clonotype to identify antigen-specific CD4⁺ T cells that integrated the biochemical properties of the TCR amino acids and the transcriptional signatures of specific cells (Fig. 3a). Model hyperparameters were empirically based using a bootstrap approach (Extended Data Fig. 3a,b). Each model in Trex showed high fidelity in the return of unique latent dimensional embeddings across sequences (Extended Data Fig. 3c) and runtimes less than 20 s for 50,000 unique TCR sequences (Extended Data Fig. 3d). The latent dimensional embeddings were based on the output of neural network-based models called variational autoencoders, which transform the amino acid sequence of each clonotype into a matrix based on Kidera factors before encoding (Fig. 3a). For a given clonotype, a centroid-like approach was used to select the best representative cell to use for RNA expression based on the minimal Euclidean distance across the calculated principal components (Fig. 3a), similar to clonotype neighbor graph analysis (CoNGA)³⁰. For a given clonotype, the TRA, TRB and RNA vectors were then coembedded, and a nonlinear dimensional reduction was calculated to represent an immune response at both the transcriptional and repertoire levels simultaneously (Fig. 3a).

We used Trex to examine all dLN CD4⁺ T_{IFN} cells (clusters c0–c11) and generated a PHATE-based manifold of the resulting

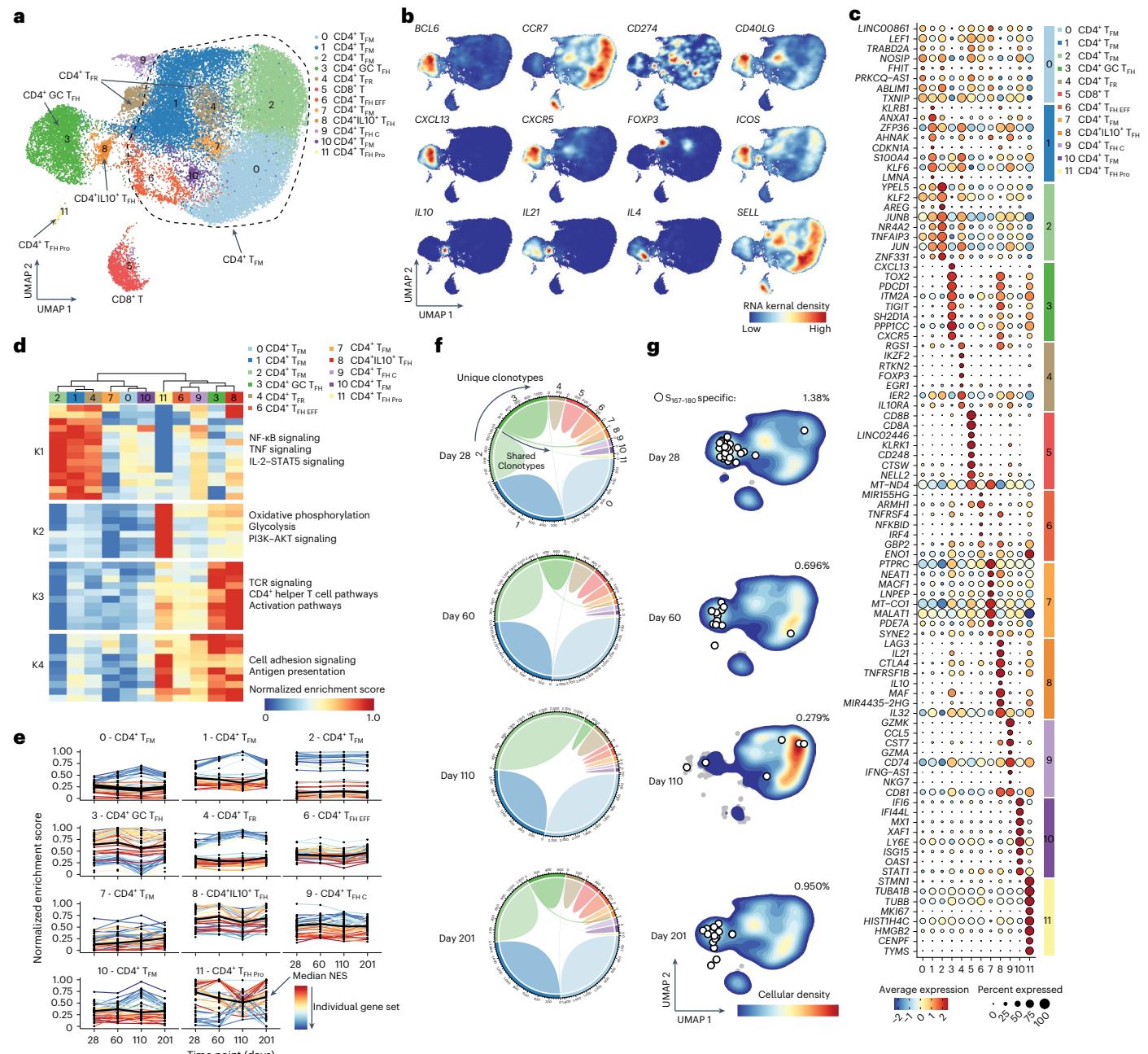


Fig. 2 | Diverse CD4⁺ T_{FH} cell and CD4⁺ T_{FM} cell transcriptional phenotypes are detected in the dLN after BNT162b2 mRNA vaccination. a, UMAP of the subset of CD4⁺ T_{FH} cells and CD4⁺ T_{FM} cells from Fig. 1b found in the dLN of 368-01a, 368-04, 368-13, 368-16, 368-20 and 368-22 on days 28, 60, 110 and 201 after the first dose of the BNT162b2 vaccine; T_{FR}, follicular T_{reg} cells. b, Gene-weighted density estimates of the indicated transcripts overlaid on the UMAP as in a. c, Top eight or fewer differentially expressed genes in clusters c0–c11 based on UMAP analysis, as in a. Dot size represents the percentage of cells expressing the gene, and color is assigned based on scaled expression value. d, Heat map of median gene set enrichment for the significantly altered gene sets in clusters c0–c4 and c6–c11, as in a. K1–K4 represent k-means clustering of gene sets with general

summaries of groupings listed to the right of each group. Significance is defined as an adjusted *P* value of <0.05 via two-way analysis of variance. e, Normalized gene set enrichment values in clusters c0–c4 and c6–c11 at days 28, 60, 110 and 201 after the first dose of the BNT162b2 vaccine. Colors indicate individual gene sets; NES, normalized enrichment score. f, Circos plot showing overlap of unique individual TCR clonotypes at days 28, 60, 110 and 201 after the first dose of the BNT162b2 vaccine, with ribbons between clusters representing overlapping clonotypes. g, Relative CD4⁺ T_{FH} and CD4⁺ T_{FM} cell densities at days 28, 60, 110 and 201 after the first dose of the BNT162b2 vaccine. S_{167–180}-specific TCRs are represented by white dots. Shown is the percentage of S_{167–180}-specific CD4⁺ T_{FH} and CD4⁺ T_{FM} cells among total CD4⁺ T_{FH} and CD4⁺ T_{FM} cells at each time point.

data that contained six independent clusters denoted Trex-C0–Trex-C5 (Fig. 3b; <https://cellpilot.emed.wustl.edu>). Transcripts of various CD4⁺ T_{FH} cell genes partitioned throughout the manifold (Fig. 3c), consistent with the inclusion of both transcriptional and TRA–TRB properties in the model. We found that previously known spike-specific TCR clonotypes colocalized into unique and

very focal areas within clusters Trex-C0, Trex-C1 and Trex-C3 of the PHATE-based manifold (Fig. 3d). TRA and TRB CDR3 in these three clusters shared related amino acid biochemical properties (Fig. 3e) that were similar to those observed in published spike-specific clonotypes. Comparison of the overlap of the nearest neighbors between the Trex- and CoNGA-derived TCR vectors

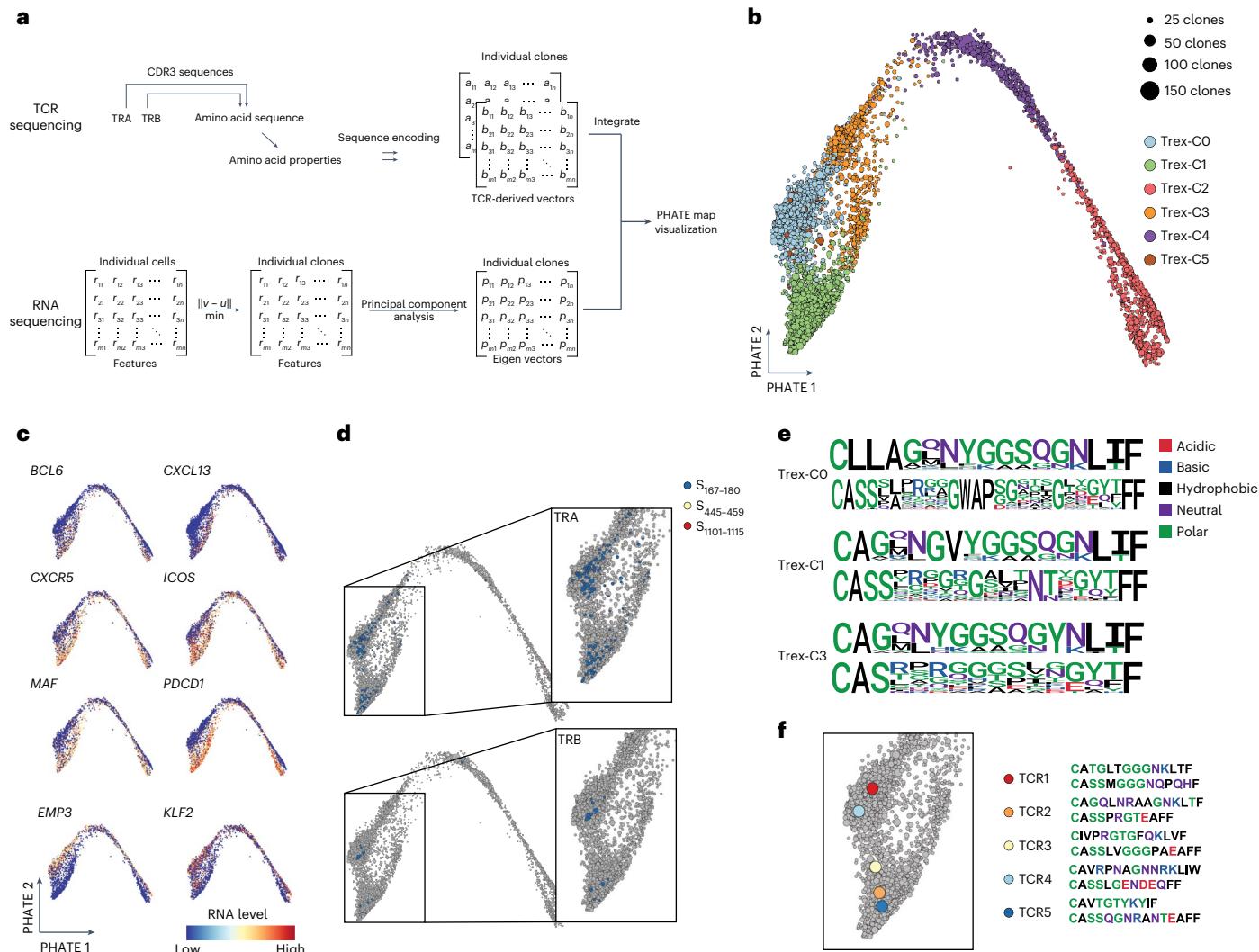


Fig. 3 | Coembedding of single-cell TCR and RNA values from dLN CD4⁺ T_{FH} and CD4⁺ T_{Fum} cells using Trex identifies spike-specific responses.

a, Graphical representation showing the computational embedding of TCR CDR3 amino acid sequences and single-cell RNA data from CD4⁺ T_{FH} cells to generate a heat diffusion-based manifold of CD4⁺ T_{FH} cells. Matrices are rescaled based on nearest neighbors, and the corrected values are then used for dimensional reduction. **b**, PHATE projection of the trimodal (RNA, TRA and TRB) embedding of dLN CD4⁺ T_{FH} cells by clonotype. The total number of each unique clonotype is represented by dot size. **c**, Representative RNA expression of CD4⁺ T_{FH} cell marker

genes (BCL6, CXCL13, CXCR5, ICOS, MAF, PDCD1, EMP3 and KLF2) overlaid onto the PHATE projection, as in **b**. **d**, Location of spike-specific^{3,21} TRA (top) and TRB (bottom) CDR3 in the PHATE projection. **e**, Alignment of TRA and TRB motifs in spike-specific TCRs^{3,21} from PHATE-defined clusters Trex-C0, Trex-C1 and Trex-C3. **f**, The location within the PHATE projection (left) and sequence (right) of five candidate spike-specific TCR clonotypes derived from PHATE-defined clusters Trex-C0 and Trex-C1 that have one TCR chain appearing in more than one donor and have not been previously described as specific for SARS-CoV-2 spike protein.

(Extended Data Fig. 4a,b) indicated distinct vectors but an overlap in the nearest neighbor clusters that contained spike-specific CD4⁺ T_{FH} cell clones (Extended Data Fig. 4c). CoNGA TCR-based clustering centralized spike-specific CD4⁺ T_{FH} cell clones into a single cluster, whereas Trex-based clusters exhibited multiple small spike-specific-predominant clusters (Extended Data Fig. 4d,e).

To test whether clonally expanded dLN CD4⁺ T_{FH} cells with at least one public TRA or TRB shared in two or more donors located within clusters Trex-C0, Trex-C1 and Trex-C3 and found in proximity to other known spike-specific CD4⁺ T cell clonotypes in the Trex PHATE-based manifold had a high probability of being spike-specific, we chose five TCR candidates that fitted these criteria (Fig. 3f) and were distributed uniquely into multiple CoNGA TCR-based clusters and Trex-based clusters (Extended Data Fig. 4f). We synthesized the five TCRs, cloned them into a retroviral transduction system³¹ and transduced them into primary human CD4⁺ T cells or Jurkat cells expressing an NFAT-green

fluorescent protein (NFAT-GFP) reporter. To determine epitope specificity, we mapped the responsiveness of each transduced TCR (TCR1–TCR5) to overlapping spike peptides in vitro. All five candidate TCRs were spike-specific (Extended Data Fig. 5). TCR2 mapped to S_{167–180} and bound the HLA-DPB1*04:01–S_{167–180} tetramer (Extended Data Fig. 5) but did not share the TRA CDR3 motif previously characterized as S_{167–180} specific³. The spike-specific epitopes for TCR3, TCR4 and TCR5 were restricted by HLA-DRB5*02:02, DRB1*07:01 and DPB1*02:01, respectively (Extended Data Fig. 6). Next, we selected all members of the TRA–TRB families with highly related TCRs to the five index TCR candidates that were experimentally determined to be spike-specific. This expanded the total number of analyzed spike-specific dLN CD4⁺ T_{FH} cells from 164 to 238 (Supplementary Table 1). Thus, Trex, along with in vitro validation of potential antigen-specific TCRs, expanded the total number of analyzed spike-specific dLN CD4⁺ T_{FH} cells in our single-cell dataset by nearly 50%.

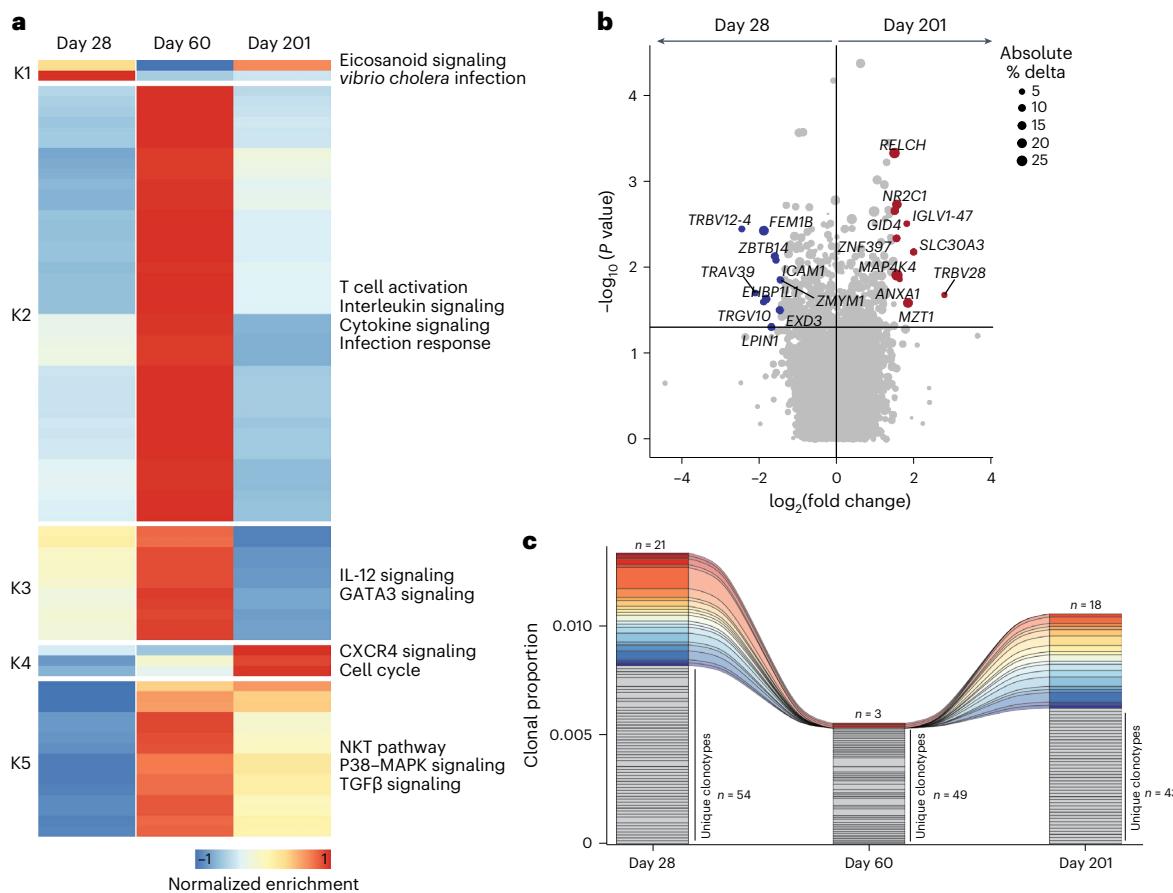


Fig. 4 | Spike-specific CD4⁺ T_{FH} cell transcriptional phenotypes detected in the dLNs change over time. **a**, Median gene set enrichment heat map showing immune-related gene sets in dLN CD4⁺ T_{FH} cells at days 28, 60 and 201 after the first dose of the BNT162b2 vaccine with five distinct clusters defined by k-means (K1–K5). **b**, Volcano plot of differential gene expression in dLN spike-specific CD4⁺ T_{FH} cells at day 28 ($n = 94$) and day 201 ($n = 70$) after the first dose of the BNT162b2 vaccine. Size is based on the change in the percentage of

cells expressing the gene at day 28 compared to at day 201. Statistical testing performed using two-sided MAST testing without adjustment for multiple comparisons. **c**, Clonal proportion of the spike-specific dLN CD4⁺ T_{FH} cell repertoire at days 28, 60 and 201 after the first dose of the BNT162b2 vaccine. The top number on the bar plot indicates the number of clones at the specific time point shared across more than one time point, whereas the number on the bottom indicates the number of spike-specific clones unique to the time point.

Spike-specific dLN CD4⁺ T_{FH} cell gene expression varies over time

We next explored the phenotypic dynamics of the expanded dataset of 238 spike-specific dLN CD4⁺ T_{FH} cells on days 28, 60 and 201 following mRNA vaccination (Fig. 4). dLN CD4⁺ T_{FH} cells from day 110 after vaccination included significantly fewer spike-specific cells (10) than cells from days 28 (94), 60 (64) and 201 (70) and were therefore excluded from the analysis. Gene set enrichment analysis revealed elevated T cell activation, interleukin signaling, cytokine signaling, infection response, IL-12 signaling, GATA3 signaling, NKT pathway genes, P38-MAPK signaling and TGF β signaling pathways at day 60, the peak of the GC response (Fig. 4a). Gene sets representing CXCR4 signaling and cell cycle progression were significantly enriched in spike-specific CD4⁺ T_{FH} cells at day 201 at the end of the GC response (Fig. 4a).

We detected several genes that were differentially expressed between day 28 and day 201, the beginning and end of the GC reaction (Fig. 4b and Supplementary Table 2). Early spike-specific dLN CD4⁺ T_{FH} cells (day 28) showed higher expression of *ICAM1* (Fig. 4b), suggesting enhanced activation and clustering of CD4⁺ T_{FH} cells during the early GC response, and higher expression of *ZBTB14* (Fig. 4b), a poorly characterized member of the zinc finger and BTB domain family of transcription factors, which also includes Bcl-6 (*ZBTB27*)³². Genes involved in cholesterol metabolism (*RELCH*), ubiquitination (*GID4*) and intracellular signaling (*MAP4K4* and *ANXA1*) were upregulated in

spike-specific CD4⁺ T_{FH} cells at day 201 (Fig. 4b). To evaluate the paired TCR clonotypes found in the dLN spike-specific CD4⁺ T_{FH} cells at various time points, we tracked 21 identical paired TCR clonotypes observed at more than one time point during the ongoing GC response (Fig. 4c). These TCR clonotypes accounted for between 5% and 28% of the total number of sequenced cells (Fig. 4c), indicating persistence or proliferation of clonally identical spike-specific CD4⁺ T_{FH} cells within the GC over time. These observations indicated that spike-specific CD4⁺ T_{FH} cells in the dLN had distinct transcriptional signatures early (day 28), at peak (day 60) and late (day 201) in the GC response and that clonally identical spike-specific CD4⁺ T_{FH} cell populations persisted throughout the course of the human GC response.

Circulating and dLN T cells show minimal overlap

To determine if clonally identical populations of spike-specific CD4⁺ T cells could be found in the blood during the ongoing GC reaction, we assessed our dataset for identical paired TRA-TRB sequences in all the sequenced dLN and blood T cells from three of the six individuals vaccinated with BNT162b2 that had matched blood and dLN samples at day 110 (368-01a, 368-13 and 368-22) and a paired blood and dLN sample from donor 368-01a at day 201 after vaccination, for which the analysis included all CD4⁺ T cells and CD8⁺ T cells. Despite 415 spike-specific CD4⁺ TCR clonotypes identified in the three donors in both the dLNs (91 spike-specific clonotypes) and blood (324 spike-specific clonotypes)

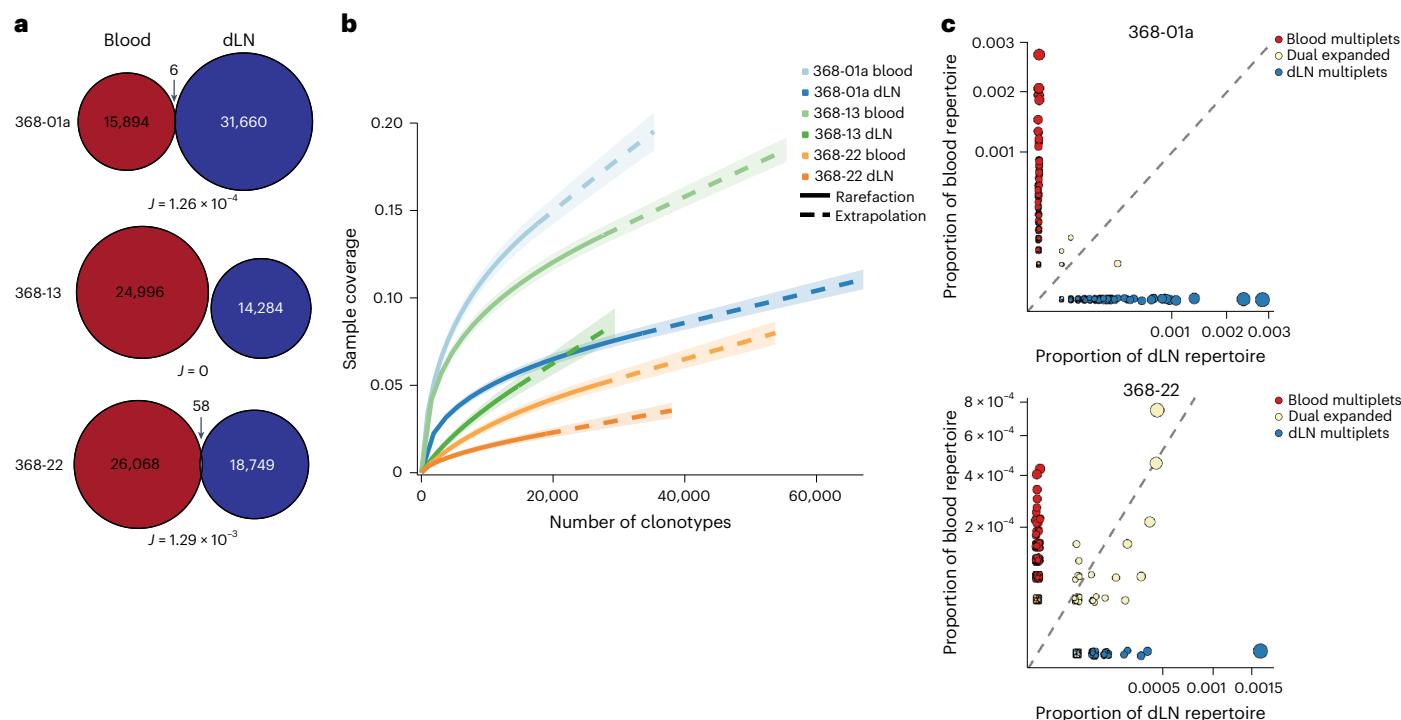


Fig. 5 | TCR sequencing revealed limited overlap in the clonal TCR repertoire between the dLN and blood 3–6 months after mRNA vaccination.

a, Representation of total T cell clonal overlap between blood and dLNs in donors 368-01a, 368-13 and 368-22 (blood samples from all three donors on days 110 and 201, dLN samples from 368-01a on days 28, 60, 110 and 201, dLN samples from 368-13 on days 60 and 110 and dLN samples from 368-22 on days 60 and 110 after the first dose of the BNT162b2 vaccine). Numbers indicate unique TCR clones,

and J is the calculated Jaccard stability index. **b**, Rarefaction and extrapolation of all TCR clones included in **a** for donors 368-01a, 368-13 and 368-22. The dotted line indicates the point of extrapolation, and the ribbon is the 95% confidence interval. **c**, Scatter plot showing the dLN or blood location and proportion of the total TCR repertoire for each TCR clone found in donors 368-01a and 368-22. Overlapping clonotypes are indicated in yellow.

at days 110 and 201, we found no clonally identical paired TCRs in the blood and dLNs in these four matched samples (data not shown). Expansion of the analysis found minimal overlap between the paired TCR repertoire in the blood and dLNs when all blood and dLN samples from these three donors (blood samples from all three donors on days 110 and 201, dLN samples from 368-01a on days 28, 60, 110 and 201, dLN samples from 368-13 on days 60 and 110 and dLN samples from 368-22 on days 60 and 110) were analyzed together (Fig. 5a). We found 6 overlapping TCRs (of 47,560 sequenced) in donor 368-01a, no overlapping TCRs (of 39,280 sequenced) in donor 368-13 and 58 overlapping TCRs (of 44,817 sequenced) in donor 368-22 (Fig. 5a). Rarefaction analysis suggested adequate sampling depth to fully represent the diversity of the TCR repertoire in both the dLN and blood compartments in these three donors (Fig. 5b), suggesting that these two compartments represented distinct populations of clonally diverse T cells 3–6 months after vaccination.

The majority of shared clonotypes between the blood and dLNs identified in donors 368-01a and 368-22 represented relatively infrequent paired TCR clonotype populations found in only one or two T cells in either the blood or the dLNs (Fig. 5c and Supplementary Tables 3 and 4) rather than clonally expanded populations, with four notable exceptions, all of which were CD8⁺ T cell populations found in 13–19 CD8⁺ T cells (Fig. 5c and Supplementary Tables 3 and 4). Based on transcriptional profiles, 48% of the populations with overlapping TCR clonotypes were CD8⁺ T cells (Supplementary Tables 3 and 4), despite the magnetic enrichment of the blood samples for CD4⁺ T cells (>97% purity before sequencing), suggesting that the frequency of overlapping blood and dLN clonal CD4⁺ T cell populations was substantially less than that observed for CD8⁺ T cells. None of the 58 overlapping TCR clonotypes in donor 368-22 contained CDR3 sequences of known

SARS-CoV-2 spike-specific CD4⁺ T cells, including those determined to be spike-specific here (Supplementary Table 4). Three of the six overlapping clonotypes in donor 368-01a were SARS-CoV-2 spike-specific CD4⁺ T cells (Supplementary Table 3), two were S_{167–180} specific and a third was S_{120–136} specific (TCR5; Supplementary Table 3). The three overlapping spike-specific CD4⁺ T cell clonotypes were found at days 28 and 60 in the dLN and at days 110 and 201 in the blood (Supplementary Table 3), perhaps indicating the emergence of memory T_{FH} cells from the dLN to the blood late in the course of the GC response. In summary, we found that 3 to 6 months following mRNA vaccination, the blood and dLNs contain distinct clonotypic populations of spike-specific CD4⁺ T cells.

Spike-specific CD4⁺ T cells are detected in the blood and dLNs

We next explored the transcriptional signatures of SARS-CoV-2 spike-specific CD4⁺ T cells found in the blood and dLN samples collected at days 110 and 201 and included both CD4⁺ T_{FH} cells and non-CD4⁺ T_{FH} cells from all sequenced dLN and blood samples at these time points, including TCRs that were S_{167–180} specific, previously published spike-specific TCRs and the five new clonotype families identified using Trex. A broad evaluation of the transcriptional differences between the blood and dLN compartments identified the upregulation of *PDCD1* and *CXCL13* in dLN spike-specific CD4⁺ T cells compared to in blood spike-specific CD4⁺ T cells at both days 110 and 201 (Fig. 6a and Supplementary Table 5). *REL* and *RELB*, which are involved in canonical and non-canonical NF-κB signaling, respectively, and *CST7*, which encodes a cysteine protease inhibitor, were significantly upregulated in blood spike-specific CD4⁺ T cells compared to in dLN spike-specific CD4⁺ T cells (Fig. 6a and Supplementary Table 5).

Gene set enrichment analysis of these data indicated substantial similarity between the spike-specific CD4⁺ T cells in the peripheral

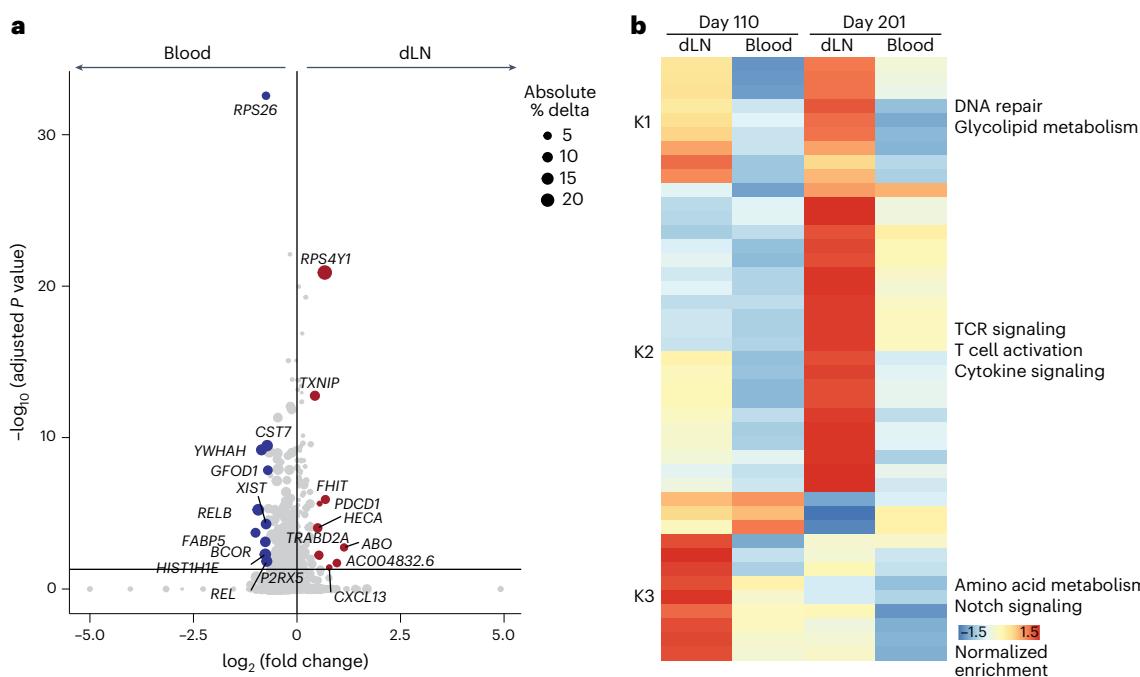


Fig. 6 | Total spike-specific blood CD4⁺ T cells are transcriptionally distinct from the total spike-specific CD4⁺ T cell population in the dLN. **a**, Volcano plot of differential gene expression between all spike-specific CD4⁺ T cells in the dLN ($n = 533$) and blood ($n = 938$). The size of the points is based on the difference in the percentage of cells expressing each gene in the dLN compared

to in the blood. Statistical testing was performed using two-sided MAST testing with adjustment for multiple comparisons. **b**, z -scaled median gene set enrichment heat map for immune-related gene sets found in dLN and blood spike-specific CD4⁺ T cells at days 110 and 201 after the first dose of the BNT162b2 vaccine.

blood and some differences between the dLN spike-specific CD4⁺ transcriptional profiles at days 110 and 201 (Fig. 6b). Spike-specific CD4⁺ T cells in the blood had less DNA repair and glycolipid metabolism signaling than the dLN samples at days 110 and 201 (Fig. 6b). We observed enrichment of TCR signaling, T cell activation and cytokine signaling pathways in the dLN samples at day 201, whereas the dLN samples at day 110 had significantly elevated amino acid metabolism and Notch signaling compared to the dLN samples at day 201 (Fig. 6b). Thus, evaluation of all spike-specific CD4⁺ T cells in both the blood and the dLNs agnostic to T_{FH} cell markers found signatures of the ongoing CD4⁺ GC T_{FH} cell response in the dLN and evidence of upregulated NF-κB signaling in spike-specific CD4⁺ T cells in the blood.

Infection induces a distinct spike-specific CD4⁺ T cell phenotype

We next leveraged our ability to detect large numbers of spike-specific CD4⁺ T cells in *HLA-DPB1*04:01*⁺ individuals to compare the transcriptional phenotype of peripheral blood spike-specific CD4⁺ T cells in BNT162b2-vaccinated individuals to the phenotype of peripheral blood spike-specific CD4⁺ T cells from individuals after acute symptomatic primary infection with SARS-CoV-2. We generated a new dataset that included single-cell RNA-seq and TCR sequencing data from peripheral blood mononuclear cells (PBMCs) collected at days 110 (3 months) and 201 (6 months) after the first dose from four BNT162b2-vaccinated individuals (three men and one woman, age range of 34–38) and magnetically enriched (>97% purity) CD3⁺ T cells from PBMCs collected 1 month, 3 months and 6 months after infection from six *HLA-DPB1*04:01*⁺ individuals hospitalized with moderate ($n = 3$, all men, age range of 53–75, WHO severity scale of 3–4) or severe ($n = 3$, all men, age range of 53–73, WHO severity scale range of 6–8, one of which died) coronavirus disease 2019 (COVID-19) during the first wave of the pandemic (April to August of 2020) before the introduction of vaccines³³ (Fig. 7a and Extended Data Tables 1 and 2). All ten donors were exposed to the SARS-CoV-2 spike

antigen for the first time, either through the two-dose BNT162b2 vaccine or through natural infection. UMAP projection of all spike-specific CD4⁺ T cells identified nine clusters: CD4⁺ central memory (T_{CM}) cells (0), T_{CM} cells expressing *SLC2A3* that encodes the glucose transporter 3 protein GLUT3 (CD4⁺GLUT3⁺ T_{CM} cells; 1), CD4⁺ effector memory T (T_{EM}) cells (2), CD4⁺GLUT3⁺ T_{EM} cells (3), a population of CD4⁺ T_{CM} cells expressing the antiapoptotic transcripts *BCL2* and *GIMAP5* (CD4⁺Bcl-2⁺ T_{CM} cells; 4), two clusters of cytotoxic CD4⁺ T_{EM} cells (T_{EM}; 5 and 8), *FOXP3*⁺CD4⁺ regulatory T (T_{reg}) cells (6) and a population of CD4⁺CD52⁺ T_{reg} cells³⁴ (7; Fig. 7b; <https://cellpilot.emed.wustl.edu>). All clusters were found in both infected and vaccinated individuals at the matched 3- to 6-month time points (Fig. 7c). CD4⁺Bcl-2⁺ T_{CM} cells, CD4⁺GLUT3⁺ T_{EM} cells and CD4⁺CD52⁺ T_{reg} cells represented a significantly higher proportion of the spike-specific CD4⁺ T cells at months 3 and 6 after antigen exposure in vaccinated compared to infected individuals (Fig. 7c). CD4⁺ T_{CM} cells and CD4⁺ T_{EM} cells were a significantly higher proportion of spike-specific CD4⁺ T cells at months 3 and 6 after antigen exposure in infected donors (Fig. 7c). Relatively small increases in the total proportion of CD4⁺ T_{CM} cells and CD4⁺ T_{EM} cells and a slight decrease in the total proportion of CD4⁺ T_{EMC} cells were detected between infected individuals with moderate or severe COVID-19 at the acute (day 18–36) time point (Extended Data Fig. 7). The two populations of CD4⁺ T_{EMC} cells (5 and 8) were characterized by high expression of cytotoxic cytokines (*CCL4* and *CCL5*) and granzymes (*GZMA*, *GZMK* and *GZMH*; Fig. 7d). We also observed distinct populations of circulating *FOXP3*⁺CD4⁺ T_{reg} cells (6) and *FOXP3*^{lo}CD52^{hi}CD4⁺ T cells (7; Fig. 7d), which were reported to suppress antigen-specific T cell responses through soluble CD52 ligation of Siglec-10 on target cells³⁴.

Expanded clonotypes of spike-specific TCRs with highly related TCRs suggestive of clonal groups were found in all clusters except the small CD4⁺ T_{EMC} cluster 8 and in every individual donor (Fig. 7e). Most of the spike-specific TCR clonal groups were found in the CD4⁺ T_{CM}, CD4⁺GLUT3⁺ T_{CM} and CD4⁺Bcl-2⁺ T_{CM} cell clusters (Fig. 7e). CD4⁺ T_{CM}

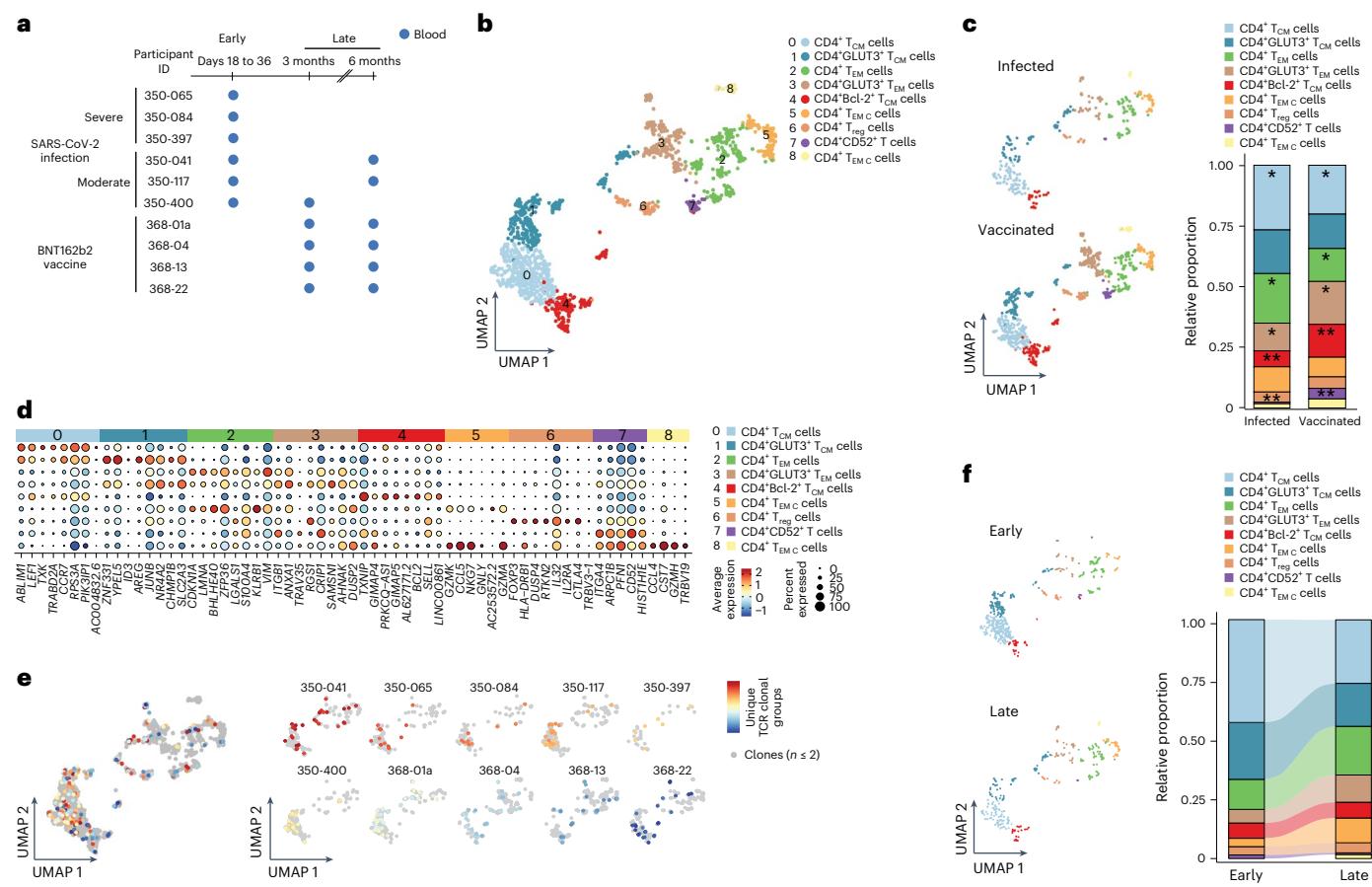


Fig. 7 | Circulating blood spike-specific CD4⁺ T cells induced by infection are transcriptionally distinct from those induced by mRNA vaccination. **a**, Schematic showing the timeline for blood sample collection 3 months and 6 months after the first dose of the BNT162b2 vaccine in donors (368-01a, 368-04, 368-13 and 368-22) or following SARS-CoV-2 infection in individuals who developed moderate (350-041, 350-117 and 350-400) or severe (350-065, 350-084 and 350-397) disease. **b**, UMAP projection of all circulating blood spike-specific CD4⁺ T cells in all donors as in **a**, including 693 cells from SARS-CoV-2-infected donors and 804 cells from BNT162b2-vaccinated donors (1,497 total cells). **c**, UMAP projection (left) and proportion breakdown (right) of cluster composition for circulating blood spike-specific CD4⁺ T cells in donors 368-01a, 368-04, 368-13, 368-22, 350-041, 350-117 and 350-400 at 3 months and 6 months after either infection with SARS-CoV-2 (infected, $n = 289$) or vaccination with the BNT162b2 vaccine (vaccinated, $n = 804$). Statistical significance was based on bootstrapping 1,000 times to form a null distribution. An asterisk (*) indicates a corrected P value of <0.05 . Two asterisks (**) indicate a corrected P value of <0.01 . **d**, Top eight or fewer cluster-defining differentially expressed genes in clusters 0–8 as in **b**. **e**, TCR cluster assignments based on normalized Levenshtein distance of the CDR3 sequence across donors. Only cluster assignments with more than two clonotypes were retained. **f**, UMAP and proportion breakdown of circulating blood spike-specific CD4⁺ T cells in infected donors 350-041, 350-117 and 350-400 at days 18 to 36 (early) and at 3–6 months (late).

368-04, 368-13, 368-22, 350-041, 350-117 and 350-400 at 3 months and 6 months after either infection with SARS-CoV-2 (infected, $n = 289$) or vaccination with the BNT162b2 vaccine (vaccinated, $n = 804$). Statistical significance was based on bootstrapping 1,000 times to form a null distribution. An asterisk (*) indicates a corrected P value of <0.05 . Two asterisks (**) indicate a corrected P value of <0.01 . **d**, Top eight or fewer cluster-defining differentially expressed genes in clusters 0–8 as in **b**. **e**, TCR cluster assignments based on normalized Levenshtein distance of the CDR3 sequence across donors. Only cluster assignments with more than two clonotypes were retained. **f**, UMAP and proportion breakdown of circulating blood spike-specific CD4⁺ T cells in infected donors 350-041, 350-117 and 350-400 at days 18 to 36 (early) and at 3–6 months (late).

cells from infected individuals comprised a larger proportion of total spike-specific CD4⁺ T cells at days 18–36 after infection than at months 3–6 after infection (Fig. 7f). CD4⁺T_{EMC} cells comprised a higher proportion of circulating spike-specific CD4⁺ T cells in infected individuals at months 3–6 after infection than at days 18–36 after infection (Fig. 7f). Thus, circulating spike-specific CD4⁺ T cells exhibit distinct phenotypes following primary induction by either infection or vaccination.

Discussion

Using previously published spike-specific TCR sequences and Trex as a tool to assist with additional reverse epitope discovery^{30,35}, here, we longitudinally tracked the evolution of large numbers of SARS-CoV-2 spike-specific CD4⁺ T cells in human blood and dLNs in the first 6 months after SARS-CoV-2 mRNA vaccination or infection. An interactive online data portal, CellPilot (<https://cellpilot.emed.wustl.edu>), allows for rapid and detailed interrogation of our dataset.

Recent supervised and unsupervised informatic tools for analyzing TCRs and antigen specificity have been developed^{36–41}. Trex is a TCR analysis platform built to combine deep variational autoencoders with gene expression data at the single-cell level. Although several

methods on the combination of TCR data and gene expression have been published^{30,40}, Trex offers up to eight variational autoencoding models (four per TCR chain) and a generative artificial intelligence approach to encode TCR amino acid sequences into latent dimensional space. In addition, the latent dimensional space of the TCRs can be used adaptively to filter and cluster or can be used as a layer input for multimodal dimensionality reduction. In the future, the use of this technique to combine single-cell RNA, protein and chromatin accessibility quantification with vectorized TCRs could allow for an even more comprehensive analysis of antigen-specific immune responses.

Despite identifying 11 CD4⁺T_{FH} and CD4⁺T_{FM} cell transcriptional phenotypes in the dLNs following vaccination, the majority of spike-specific CD4⁺T_{FH} cells exhibited the classical CD4⁺GC T_{FH}^{4,22} and CD4⁺IL-10⁺T_{FH}^{15,23} phenotypes. We identified a large number of overlapping paired TCR clonotypes between CD4⁺GC T_{FH} cells and CD4⁺IL-10⁺T_{FH} cells between days 28 and 201 after vaccination, suggesting a common origin of these two effector T_{FH} cell populations, despite significant transcriptional differences between the two subsets, which implied very different functional roles. Differential IL21 and IL4 expression in CD4⁺GC T_{FH} and CD4⁺IL-10⁺T_{FH} cells was reminiscent of

the segregation of these important functional cytokines in time and space within the mouse GC after infection^{42,43}.

We did not observe substantial overlap between paired human spike-specific CD4⁺ TCR sequences found in the dLNs and those found in the matched blood samples at days 110 and 201 during the late GC response, in line with a report that found minimal clonal overlap between paired TCRs from CD4⁺ T cells in the blood and LN compartments samples from deceased organ donors, but did detect overlap between CD8⁺ T cells in the blood and LN compartments⁴⁴. Indeed, many of the overlapping clonotypes that we identified in the dLNs and blood in this study were contaminating CD8⁺ T cells. The three overlapping spike-specific CD4⁺ TCRs that we discovered in one donor were found at days 28 and 60 in the dLNs and at days 110 and 201 in the blood. These cells may represent the first emergence of circulating CD4⁺ T_{FH} cells from the dLNs.

Our tracking and transcriptional phenotyping of large numbers of spike-specific CD4⁺ T cells allowed us to gain important insights into the execution of the spike-specific CD4⁺ T_{FH} cell response in the dLNs after mRNA vaccination. Peak T cell activation, interleukin signaling and cytokine signaling transcriptional activity were observed at day 60. We found upregulation of the expression of genes associated with CXCR4 signaling at day 201, near the end of the GC response, raising the possibility that this pathway may play a role in the termination of the GC response in humans. Notably, CXCR4 signaling is critical in localizing CD4⁺ T cells found within the GC to the dark zone⁴⁵. We speculate that the localization of antigen-specific CD4⁺ GC T_{FH} cells in the dark zone toward the end of the GC response may facilitate the termination and collapse of the ongoing GC response.

Finally, our comparison of the mRNA vaccine-induced spike-specific CD4⁺ T cells in the blood to the spike-specific CD4⁺ T cells from the blood of SARS-CoV-2-infected individuals identified a significantly higher proportion of spike-specific Bcl-2⁺ CD4⁺ T_{CM} cells with higher expression of a prosurvival/antiapoptotic transcriptional program in vaccinated individuals. Therefore, unique long-term transcriptional profiles were induced in spike-specific memory CD4⁺ T cells depending on the context of initial antigen exposure (infection or mRNA vaccination).

Our work has limitations. The present study evaluated CD4⁺ T cell responses from six mRNA-vaccinated and six SARS-CoV-2-infected donors. Although our results are reproducible across this cohort, their broad applicability across larger populations of individuals cannot be adjudicated at this time. Our focus on *HLA-DPB1*04⁺* individuals, while necessary to obtain sufficient numbers of antigen-specific cells for the unique analyses that we performed, may have introduced unrecognized bias into our results, and further validation of our findings would be required to ensure that these findings apply to individuals without this *HLA* allele. In summary, we developed a broad single-cell transcriptional atlas of human spike-specific CD4⁺ T cells in the blood and dLNs in the first 6 months following primary exposure to SARS-CoV-2 spike protein through BNT162b2 mRNA vaccination or infection.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-024-01888-9>.

References

- Anderson, E. J. et al. Safety and immunogenicity of SARS-CoV-2 mRNA-1273 vaccine in older adults. *N. Engl. J. Med.* **383**, 2427–2438 (2020).
- Painter, M. M. et al. Rapid induction of antigen-specific CD4⁺ T cells is associated with coordinated humoral and cellular immunity to SARS-CoV-2 mRNA vaccination. *Immunity* **54**, 2133–2142 (2021).
- Mudd, P. A. et al. SARS-CoV-2 mRNA vaccination elicits a robust and persistent T follicular helper cell response in humans. *Cell* **185**, 603–613 (2022).
- Crotty, S. Follicular helper CD4 T cells (T_{FH}). *Annu. Rev. Immunol.* **29**, 621–663 (2011).
- Vinuesa, C. G., Linterman, M. A., Yu, D. & MacLennan, I. C. M. Follicular helper T cells. *Annu. Rev. Immunol.* **34**, 335–368 (2016).
- Crotty, S. T follicular helper cell biology: a decade of discovery and diseases. *Immunity* **50**, 1132–1148 (2019).
- Nurieva, R. I. et al. Bcl6 mediates the development of T follicular helper cells. *Science* **325**, 1001–1005 (2009).
- Johnston, R. J. et al. Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325**, 1006–1010 (2009).
- Yu, D. et al. The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity* **31**, 457–468 (2009).
- Akiba, H. et al. The role of ICOS in the CXCR5⁺ follicular B helper T cell maintenance in vivo. *J. Immunol.* **175**, 2340–2348 (2005).
- Crotty, S., Kersh, E. N., Cannons, J., Schwartzberg, P. L. & Ahmed, R. SAP is required for generating long-term humoral immunity. *Nature* **421**, 282–287 (2003).
- Brenna, E. et al. CD4⁺ T follicular helper cells in human tonsils and blood are clonally convergent but divergent from non-T_{FH} CD4⁺ cells. *Cell Rep.* **30**, 137–152 (2020).
- Dan, J. M. et al. Recurrent group A Streptococcus tonsillitis is an immunosusceptibility disease involving antibody deficiency and aberrant T_{FH} cells. *Sci. Transl. Med.* **11**, eaau3776 (2019).
- Heit, A. et al. Vaccination establishes clonal relatives of germinal center T cells in the blood of humans. *J. Exp. Med.* **214**, 2139–2152 (2017).
- Cañete, P. F. et al. Regulatory roles of IL-10-producing human follicular T cells. *J. Exp. Med.* **216**, 1843–1856 (2019).
- Del Alcazar, D. et al. Mapping the lineage relationship between CXCR5⁺ and CXCR5⁻ CD4⁺ T cells in HIV-infected human lymph nodes. *Cell Rep.* **28**, 3047–3060 (2019).
- Poon, M. M. L. et al. SARS-CoV-2 infection generates tissue-localized immunological memory in humans. *Sci. Immunol.* **6**, eabl9105 (2021).
- Turner, J. S. et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature* **596**, 109–113 (2021).
- Kim, W. et al. Germinal centre-driven maturation of B cell response to mRNA vaccination. *Nature* **604**, 141–145 (2022).
- Dykema, A. G. et al. Functional characterization of CD4⁺ T cell receptors crossreactive for SARS-CoV-2 and endemic coronaviruses. *J. Clin. Invest.* **131**, e146922 (2021).
- Goncharov, M. et al. VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2. *Nat. Methods* **19**, 1017–1019 (2022).
- Ramiscal, R. R. & Vinuesa, C. G. T-cell subsets in the germinal center. *Immunol. Rev.* **252**, 146–155 (2013).
- Kumar, S. et al. Developmental bifurcation of human T follicular regulatory cells. *Sci. Immunol.* **6**, eabd8411 (2021).
- Chen, M., Wang, F., Xia, H. & Yao, S. MicroRNA-155: regulation of immune cells in sepsis. *Mediators Inflamm.* **2021**, 8874854 (2021).
- Niu, L. et al. A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation. *Sci. Adv.* **6**, eaaz2059 (2020).
- Carlson, C. M. et al. Krüppel-like factor 2 regulates thymocyte and T-cell migration. *Nature* **442**, 299–302 (2006).
- Cao, Z., Sun, X., Icli, B., Wara, A. K. & Feinberg, M. W. Role of Krüppel-like factors in leukocyte development, function, and disease. *Blood* **116**, 4404–4414 (2010).
- Lynn, R. C. et al. c-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature* **576**, 293–300 (2019).

29. Koizumi, S. et al. JunB regulates homeostasis and suppressive functions of effector regulatory T cells. *Nat. Commun.* **9**, 5344 (2018).
30. Schattgen, S. A. et al. Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). *Nat. Biotechnol.* **40**, 54–63 (2022).
31. Abdulhaqq, S. et al. Identification and characterization of antigen-specific CD8⁺ T cells using surface-trapped TNF- α and single-cell sequencing. *J. Immunol.* **207**, 2913–2921 (2021).
32. Cheng, Z.-Y., He, T.-T., Gao, X.-M., Zhao, Y. & Wang, J. ZBTB transcription factors: key regulators of the development, differentiation and effector function of T cells. *Front. Immunol.* **12**, 713294 (2021).
33. Mudd, P. A. et al. Distinct inflammatory profiles distinguish COVID-19 from influenza with limited contributions from cytokine storm. *Sci. Adv.* **6**, eabe3024 (2020).
34. Bandala-Sanchez, E. et al. T cell regulation mediated by interaction of soluble CD52 with the inhibitory receptor Siglec-10. *Nat. Immunol.* **14**, 741–748 (2013).
35. Huang, H., Wang, C., Rubelt, F., Scriba, T. J. & Davis, M. M. Analyzing the *Mycobacterium tuberculosis* immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat. Biotechnol.* **38**, 1194–1202 (2020).
36. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).
37. Mayer-Blackwell, K. et al. TCR meta-clonotypes for biomarker discovery with tcrdist3 enabled identification of public, HLA-restricted clusters of SARS-CoV-2 TCRs. *eLife* **10**, e68605 (2021).
38. Zhang, H., Zhan, X. & Li, B. GIANA allows computationally-efficient TCR clustering and multi-disease repertoire classification by isometric transformation. *Nat. Commun.* **12**, 4699 (2021).
39. Zhao, Y. et al. DeepAIR: a deep learning framework for effective integration of sequence and 3D structure to enable adaptive immune receptor analysis. *Sci. Adv.* **9**, eabo5128 (2023).
40. Zhang, Z., Xiong, D., Wang, X., Liu, H. & Wang, T. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat. Methods* **18**, 92–99 (2021).
41. Sidhom, J.-W., Larman, H. B., Pardoll, D. M. & Baras, A. S. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat. Commun.* **12**, 1605 (2021).
42. Weinstein, J. S. et al. T_{FH} cells progressively differentiate to regulate the germinal center response. *Nat. Immunol.* **17**, 1197–1205 (2016).
43. Shulman, Z. et al. Dynamic signaling by T follicular helper cells during germinal center B cell selection. *Science* **345**, 1058–1062 (2014).
44. Poon, M. M. L. et al. Tissue adaptation and clonal segregation of human memory T cells in barrier sites. *Nat. Immunol.* **24**, 309–319 (2023).
45. Silva-Cayetano, A. et al. Spatial dysregulation of T follicular helper cells impairs vaccine responses in aging. *Nat. Immunol.* **24**, 1124–1137 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2024

¹Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO, USA. ²Department of Microbiology, Korea University College of Medicine, Seoul, Korea. ³Division of Infectious Diseases, Department of Pediatrics, Washington University School of Medicine, Saint Louis, MO, USA. ⁴Department of Emergency Medicine, Washington University School of Medicine, Saint Louis, MO, USA. ⁵Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN, USA. ⁶Clinical Trials Unit, Washington University School of Medicine, Saint Louis, MO, USA. ⁷Mallinckrodt Institute of Radiology, Washington University School of Medicine, Saint Louis, MO, USA. ⁸Division of Biostatistics, Washington University School of Medicine, Saint Louis, MO, USA. ⁹Division of Infectious Diseases, Department of Internal Medicine, Washington University School of Medicine, Saint Louis, MO, USA. ¹⁰Center for Vaccines and Immunity to Microbial Pathogens, Washington University School of Medicine, Saint Louis, MO, USA. ¹¹The Andrew M. and Jane M. Bursky Center for Human Immunology and Immunotherapy Programs, Washington University School of Medicine, Saint Louis, MO, USA.  e-mail: ellebedy@wustl.edu; pmudd@wustl.edu

Methods

Participants

We included samples from two prospective observational human cohorts. Demographics and *HLA* typing of all included participants are reported in Extended Data Tables 1 and 2. In cohort 1, participants who received the primary two-dose BNT162b2 mRNA vaccine series were prospectively enrolled into observational study WU-368, approved by the Washington University in St. Louis Institutional Review Board (approval number 2020-12-081). Complete details of the study cohort have been previously published^{3,18,19}. Briefly, the cohort included 43 vaccinated individuals who provided blood samples, with an age range from 28 to 73 years; 21 were women. Fifteen participants from the same cohort (age range of 28–52 years, seven women) also provided one or more dLN FNA samples. Written informed consent was obtained from each participant.

Participants in cohort 2 were infected with SARS-CoV-2 during the first wave of the COVID-19 pandemic (April to August of 2020). Participants with acute symptomatic viral respiratory illness evaluated at Barnes-Jewish Hospital, Saint Louis Children's Hospital, Christian Hospital or affiliated Barnes-Jewish Hospital testing sites, all located in Saint Louis, MO, were enrolled into a prospective observational cohort study, WU-350. The WU-350 study was approved by the Washington University in St. Louis Institutional Review Board (approval number 2020-03-085). Full details of the cohort and inclusion criteria have been previously published³³. The six donors (age range of 53–75 years, all men, WHO severity scores of 3–8; Extended Data Table 1) included in the present manuscript tested positive for SARS-CoV-2 by a clinical PCR test. Informed consent was obtained from each participant or their legally authorized representative.

Sample preparation

Vaccinated participants underwent dLN FNA sampling as previously described⁴⁶. Briefly, draining dominant lateral axillary LNs ipsilateral to the deltoid muscle mRNA vaccination site were localized with ultrasound and sampled 28, 60, 110 and/or 201 days after the first vaccine dose with multiple passes of six separate 25-gauge needles using real-time ultrasound guidance. Each needle was flushed with 3 ml of R10 (RPMI 1640 medium containing L-glutamine supplemented with 10% fetal bovine serum (FBS) and 100 U ml⁻¹ penicillin–streptomycin), followed by three 1-ml rinses with R10. Any contaminating red blood cells (RBCs) were lysed with ACK hypotonic lysis buffer, dLN FNA cells were washed twice with P2 (1× PBS supplemented with 2% FBS and 2 mM EDTA), and cells were counted and cryopreserved in 90% FBS with 10% DMSO before storing in liquid nitrogen until analysis. Matched blood samples from vaccinated individuals 110 or 201 days after the first vaccine dose were obtained in EDTA-anticoagulated tubes, and PBMCs were prepared using Ficoll density gradient centrifugation. Contaminating RBCs were removed from PBMCs via hypotonic lysis, and PBMCs were washed, counted and cryopreserved in 90% FBS/10% DMSO and stored in liquid nitrogen until analysis. Blood samples from infected participants were collected 18–36 days and 3 or 6 months after the onset of viral respiratory illness symptoms into EDTA-anticoagulated tubes, and PBMCs were prepared using Ficoll density gradient centrifugation. Contaminating RBCs were removed from PBMCs via hypotonic lysis, and PBMCs were washed, counted and cryopreserved in 90% FBS/10% DMSO and stored in liquid nitrogen until analysis.

HLA typing

Vaccinated individuals were *HLA* typed by nanopore sequencing⁴⁷. Genomic DNA was purified using an AllPrep DNA/RNA kit (Qiagen). Target *HLA* genes were amplified by long-range PCR (NGS LR kit, One Lambda) and sequenced following the SQK-LSK109 protocol on R10.3 MinION flow cells (Oxford Nanopore Technologies). High-resolution *HLA* typing was assigned using the Athlon2 program.

For *HLA* typing of infected individuals, we extracted DNA from PBMCs using Zymo Quick-DNA Plus kits for use in the AlITYpe NGS

11-Loci Amplification kit (One Lambda). *HLA* libraries were sequenced at 150×150 base pairs (MiSeq, Illumina), and the data were analyzed with TypeStream Visual (v3.0; One Lambda).

dLN single-cell RNA-seq library preparation and sequencing

dLN FNA samples were thawed, washed with P2 and resuspended in P2. Chromium Single Cell 5' Gene Expression Dual Index libraries and Chromium Single Cell V(D)J Dual Index libraries (10x Genomics) were prepared according to the manufacturer's instructions without modifications. Both gene expression and V(D)J libraries were sequenced on a Novaseq S4 (Illumina) instrument, targeting a sequencing depth of 50,000 and 5,000 read pairs per cell, respectively.

T cell enrichment of PBMC populations for single-cell RNA-seq

Frozen PBMC samples were thawed, washed once with R10 and washed with P2. PBMCs were counted on a Cellometer Auto 2000 (Nexcelom) and resuspended to a final concentration of approximately 10⁸ cells per ml in P2. Total untouched CD3⁺ or positively selected CD4⁺ T cells were enriched using either an EasySep Human T Cell Isolation kit or an EasySep Human CD4 Positive Selection kit II, respectively, with an EasyEights magnet (STEMCELL Technologies), all per the manufacturer's instructions. Following enrichment, T cell populations were washed with P2, recounted and resuspended in PBS supplemented with 0.05% bovine serum albumin. Chromium Single Cell 5' v2 Gene Expression and Chromium Single Cell V(D)J libraries (10x Genomics) were prepared according to the manufacturer's instructions without modifications. Gene expression and V(D)J libraries were sequenced on a Novaseq S4 (Illumina) instrument.

The remaining T cells were stained for flow cytometry to verify T cell enrichment. Enriched T cells were added to a round-bottom 96-well plate and washed twice in P2. A master mix was added to the cells with the following reagents for 20 min at 4 °C: CD3 APC Fire 810 (HIT3a, Biolegend), CD4 PerCP (OKT4, Biolegend, to avoid blocking from positive selection), CD8 BV421 (RPA-T8, Biolegend), CD16 BV570 (3G8, Biolegend), CD14 APC (M5E2, Biolegend), CD19 BV750 (HIB19, Biolegend), Zombie NIR (Biolegend) diluted in Brilliant Staining buffer (50 µl per test, BD Horizon) and P2. Following staining, cells were washed three times in P2 and fixed with 1% paraformaldehyde (Electron Microscopy Sciences) for 20 min at 4 °C. Cells were washed once in P2, resuspended in P2 and stored at 4 °C until analysis within 24–48 h. Flow cytometry samples were run on an Aurora spectral flow cytometer using SpectroFlo v.2.2 software (Cytek). Flow cytometry data were analyzed using FlowJo v.10 (Treestar).

Single-cell RNA-seq processing and analysis

Filtered outputs of 10x Cell Ranger count and V(D)J pipelines were imported into R (v4.1) using the Seurat (v4.1.0) R package⁴⁸. Filtering was applied on a sequencing run basis to remove cells with less than 100 features, more than 2.5-fold the standard deviation of feature numbers and greater than 15% mitochondrial gene percentage. Doublets were estimated using the scDblFinder (v1.6.0) R package⁴⁹. Individual cells were annotated using the ProjecTILs (v2.0.3)^{50,51} and SingleR (v1.6.1) R packages⁵² and the DICE annotation dataset⁵³. Clonotypes were added to the integrated Seurat object using the scRepertoire (v1.7.0) R package⁵⁴. T cells were isolated based on the assignment of CD4/CD8 T cell annotation from ProjecTIL and the presence of a productive clonotype. Overall T cell dimensional reduction used 2,000 variable genes with the TCR genes removed to prevent bias in the manifold by clonality. The Harmony (v0.1.0) R package⁵⁵ was used in integrating multiple sequencing runs and generating the UMAP (dimensions = 1:15, epochs = 500) and clusters (resolution = 0.8, dimensions = 1:15, algorithm = 3). T_H cell UMAP embedding and clustering used dimensions of 1:20 and a resolution of 0.5. CD8⁺ T cell designations were based on the examination of the distribution of CD8 expression, and a cutoff was set for CD8A ≥ 0.4. Spike-specific cells from vaccinated and infected donor peripheral

blood were integrated using the Harmony R package, the individual sequencing run as the variable and dimensions = 1:30 with calculating UMAP (dimensions = 1:25, epochs = 500) and cluster (resolution = 0.5 and algorithm = 3). Gene expression UMAP overlays used the Nebulosa (v1.6.0) R package⁵⁰. Gene set enrichment analysis was performed using the escape (v1.4.2) R package⁵⁷ with the UCell approach⁵⁸ and the Hallmark, KEGG and BioCarta gene set libraries from GSEA⁵⁹. TCR rarefaction and extrapolation was performed using the iNEXT (v3.0.0) R package⁶⁰ using the abundance of combined TRA and TRB clonotypes by participant and tissue and default settings in terms of bootstraps, knots and Hill numbers. TCR clustering was performed using the scRepertoire package and the cluster.TCR function with the normalized edit distance threshold set to 0.85.

TCR sequencing analysis and visualization

Spike-specific clonotype annotations were assigned for both TRA and TRB and derived from previously published data^{3,20} and the VDJdb database²¹. TCR sequencing motifs were created with the msa (v1.28.0) R package set to protein alignment with the ClustalW algorithm and a max iteration of 30. The resulting aligned sequences were converted into seginer format and plotted with the ggseqlogo (v0.1) R package. Single-clonotype representation for single-cell analysis was performed similar to the previously described CoNGA³⁰. For a given combined TRA and TRB, a single transcriptome was selected based on the minimal Euclidean distance across all cells in the individual clonotype. Vectors for TRA and TRB were calculated using the TCR autoencoder Trex (v0.99.7) R package translating the CDR3 amino acid sequence into a matrix based on the Kidera factors⁶¹. For the resulting RNA principal components and embedded TCR values, the first 15 dimensions were selected and rescaled using the mutual nearest neighbor approach with $k = 100$ with the mumosa (v1.4.0) R package. The resulting values were then subjected to the PHATE algorithm with default settings with the PhateR (v1.0.7) R package⁶². Clustering was performed by generating a k -nearest neighbor igraph with the bluster (v1.6.0) R package, and clusters were calculated using the Leiden algorithm from the leidenAlg (v1.0.3) R package with a resolution of 0.7 and the number of iterations set to 5. Putative spike-specific TCRs were derived from clusters where previously identified spike-specific TCRs were present. In addition, the putative TCRs were selected based on the fact that they were clonally expanded and they expressed either an α - or β -chain that appeared in two or more donors and had not been previously shown to bind the spike epitope. Related putative spike-specific clones were called by identifying TRA or TRB CDR3 sequences within a Levenshtein distance of two and shared V genes.

Development of Trex autoencoding models

TCR embedding used training variational autoencoders on TRA and TRB CDR3 amino acid sequences, taking the AF, KF or both converted numeric matrices with 0 padding to set a CDR3 length of 60. The matrices were transformed into a one-dimensional array, and values were normalized across all sequences. Values with no variation were transformed into 0 s. Alternatively, a one-hot autoencoding approach was also trained by converting the amino acid sequence to a matrix based on the individual amino acids along the sequence. A stacked autoencoder approach was used similar to the previously described method⁴⁰ with a 128–64–30–64–128 neuron structure. The bottleneck layer consisted of a 30-neuron/vector embedding. Each autoencoder model was trained using the keras (v2.4.0) R package across 288,043 unique CDR3 amino acid TRA and 453,111 unique CDR3 amino acid TRB sequences across 15 single-cell datasets and four curated TCR databases (McPAS-TCR⁶³, VDJdb⁶⁴, IEDB⁶⁵ and PIRD⁶⁶), resulting in eight models: TRA-AF, TRA-KF, TRA-both, TRA-OHE, TRB-AF, TRB-KF, TRB-both and TRB-OHE. The models were trained using a data split of 80:20, and hyperparameters were selected based on minimal Kullback–Leibler divergence value with a batch size of 128, learning rate of 0.001 and

optimization using root mean square propagation. The TCR models and corresponding R package to run the embeddings with a Seurat or single-cell experiment object are available at <https://github.com/neborcherding/Trex>.

Putative spike-specific TCR transductants

Putative spike-specific TRA and TRB variable regions were combined in silico with mouse constant regions (mouse TRAC and mouse TRBC2) modified to include additional cysteine residues in place of serine at position 57 in mouse TRBC2 and threonine at position 47 in mouse TRAC. Using mouse constant regions prevents pairing with endogenous human TCRs following retroviral transduction of primary human T cells. The additional cysteine residues enhance α/β constant region binding affinity, increasing chimeric human variable/mouse constant TCR surface expression. Constructs containing the modified TRA and TRB were separated by a T2A sequence and synthesized to include NotI and EcoRI restriction sites at the 5' and 3' ends of the region of interest, respectively (GenScript). Synthesized constructs from GenScript were double digested with NotI and EcoRI and cloned into the pMP71 retroviral vector³¹; ligation was confirmed via sequencing of the recombinant plasmid. Recombinant pMP71 was used to transfect the 293Vec-RD114 retroviral packaging cell line (provided by BioVec Pharma) with TransIT-LT1 (Mirus Bio) transfection reagent using the manufacturer's protocol and recommended conditions. Transfection medium was removed after 24 h and replaced with fresh medium, and retrovirus-containing supernatants were collected 24 h later. Retroviral supernatants were stored at -80 °C until use.

Human CD4⁺ T cells were enriched from cryopreserved PBMCs using an EasySep Human CD4 Positive Selection kit II (STEMCELL Technologies). Isolated T cells were cultured in R10-500 (R10 supplemented with 500 U ml⁻¹ recombinant human IL-2 (BioLegend)) at 37 °C with 5% CO₂ and activated with a Miltenyi Biotec Human T Cell Activation/Expansion kit according to the manufacturer's instructions. Two days after activation/expansion, activated T cells were purified from dead cell debris and activation beads with a Ficoll gradient. Cells were washed in R10, resuspended at 2 × 10⁶ per ml in R10-500 and plated on 24-well flat-bottom tissue culture plates.

TCR RD114 retroviral supernatants were thawed, layered on top of a 20% sucrose (wt/vol) gradient and centrifuged in a microcentrifuge at 20,000g at 4 °C for 1 h. The supernatant was discarded, and the residual volume, including the retroviral pellet, was incubated with ViroMag beads (OZ Biosciences) for 15 min at room temperature. Retrovirus/beads were then added to the activated T cells in a 24-well plate, and the plate was briefly centrifuged at 1,600g for 1 min before being placed on a prewarmed magnet (OZ Biosciences) and incubated at 37 °C with 5% CO₂ for 15 min. Transduced T cells were cultured for at least 1 week before analysis with changes of R10-500 medium as needed.

Intracellular cytokine staining mapping of human TCR transductants

In total, 250,000 to 500,000 transduced CD4⁺ T cells, a portion of which were confirmed to express the recombinant chimeric TCR using a mouse TCR β -chain-specific monoclonal antibody (BV510, clone H57-597, BioLegend), were cocultured with 100,000 Epstein–Barr virus-transformed B cells from the experimental participant who expressed the index paired putative spike-specific TCR in the presence of various mapping pools of SARS-CoV-2 spike overlapping 17-mer peptides (NR-52402, BEI Resources). Each peptide was incubated at a final concentration of 1 μ g ml⁻¹. Separate unstimulated control wells with equivalent concentrations of DMSO to the final concentration of DMSO found in the peptide-stimulated condition were included. Positive control phorbol 12-myristate 13-acetate (InvivoGen) and ionomycin (InvivoGen) were added to separate wells. Cells in all conditions were cocultured in R10 medium supplemented with co-stimulatory antibodies to CD28 and CD49d (BD Biosciences). Samples with the

appropriate stimulus were incubated for 1.5 h before the addition of Brefeldin A and monensin (both from BD Biosciences) and incubated for an additional 12–16 h. Surface staining was performed, followed by fixation in 1% paraformaldehyde, permeabilization with washing buffer supplemented with 0.1% (wt/vol) saponin (Sigma) and intracellular staining using fluorescently labeled antibodies directed to cytokine antigens. We used the following antibodies: CD3 PE-Cy7 (clone UCHT1, BioLegend), CD4 APC-Cy7 (clone SK3, BioLegend), mouse TCR β -chain BV510 (clone H57-597, BioLegend), CD69 BV711 (clone FN50, BioLegend), IFN γ PE (clone B27, BioLegend), TNF α PerCP-Cy5.5 (clone MAb11, BioLegend) and IL-2 APC (clone 5344.111, BD Biosciences). The panel included Zombie NIR viability stain (BioLegend). All antibodies were used at pretitrated optimal staining concentrations. In a separate experiment performed on unstimulated TCR2 transductants, we performed the surface stain portion of the panel following incubation with an S_{167–180} HLA-DPB1*04:01 PE-labeled tetramer reagent (Washington University in Saint Louis tetramer core facility) for 15 min to confirm intracellular cytokine staining results. All samples were acquired on a Cytek Aurora spectral flow cytometer, and unmixed files were analyzed using FlowJo software (version 10, BD Biosciences). The final analysis was gated on live CD4 $^+$ T cells positive for mouse TCR β -chain.

HLA restriction determination

Human Jurkat clone E6-1 T cells were obtained from ATCC and transduced with an NFAT eGFP reporter lentivirus (BPS Bioscience), according to the manufacturer's instructions, and selected with puromycin at a final concentration of 1 $\mu\text{g ml}^{-1}$ for 7 days. NFAT–GFP reporter Jurkat T cells were then transduced with each candidate TCR retrovirus following sucrose purification, as described above. TCR-transduced NFAT–GFP reporter Jurkat T cell lines were purified by fluorescence-activated cell sorting after staining with CD3 PE-Cy7 (clone UCHT1, BioLegend), CD4 APC-Cy7 (clone SK3, BioLegend) and mouse TCR β -chain BV510 (clone H57-597, BioLegend) on a Bigfoot Spectral Cell Sorter (Thermo Fisher).

Single class II *HLA* allele-expressing antigen-presenting cells were developed by gene synthesis of select class II *HLA* alleles found in the participants from which the candidate TCRs were selected. We synthesized class II *HLA* α - and β -chains separated by a T2A sequence and included NotI and EcoRI restriction sites at the 5' and 3' ends (GeneScript). Synthesized constructs were cloned into the pMP71 retroviral transduction system as described above and transfected into the 293Vec-RD114 retroviral packaging cell line, and resulting retroviruses were used to transduce a K562-based artificial APC cell line expressing exogenous human CD64, CD80, CD83, CD74 and HLA-DM.

NFAT–GFP reporter Jurkat TCR cell lines were cocultured in R10 with individual K562 artificial APC cell lines either expressing unique class II *HLA* alleles or not expressing class II *HLA* (not *HLA* virus transduced) in the presence or absence of 10 $\mu\text{g ml}^{-1}$ of the unique SARS-CoV-2 spike peptide previously mapped to each responding TCR for a total of 16 h at 37 °C and 5% CO₂. Following coculture, cells were surface stained with CD3 PE-Cy7 (clone UCHT1, BioLegend), mouse TCR β -chain BV510 (clone H57-597, BioLegend) and Zombie NIR viability stain (BioLegend) before analysis on a Cytek Aurora spectral flow cytometer. Unmixed files were analyzed using FlowJo software (version 10, BD Biosciences).

Statistics and reproducibility

Heat maps of gene sets were derived from the intersection of significant enrichment comparisons (Bonferroni-adjusted *P* value of <0.05) by analysis of variance, Kruskal–Wallis *H*-test for multiple comparisons and *t*-test and Wilcoxon rank-sum test for binarized comparisons. For none-rank-based significance testing, distributions were evaluated before applying testing. Differential gene expression used MAST⁶⁷ using the donor as a latent variable and a pseudocount of 0.1. Cluster proportion comparisons between antigen-specific T cells used the scProportionTest (v0.0.0.9) R package with 1,000 permutations. Code

for the entire analysis is available at https://github.com/ncborcherding/COVID_TCR. The cohort sample size was based on the voluntary enrollment of participants who consented to the respective procedures in the trial, and randomization or blinding is not applicable to this study.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Single-cell data presented in the manuscript can be accessed at the publicly available Gene Expression Omnibus database under accession numbers GSE195673 and GSE249313. Datasets are also available on the interactive online tool CellPilot (<https://cellpilot.emed.wustl.edu>). CellPilot was adapted from the cellcuratoR (v0.1.0) R package⁶⁸. Single-cell data are also available at the public Zenodo repository at <https://doi.org/10.5281/zenodo.11395445> (ref. 69).

Code availability

Analysis code is available at the public Zenodo repository at <https://doi.org/10.5281/zenodo.11395445> (ref. 69).

References

46. Turner, J. S. et al. Human germinal centres engage memory and naïve B cells after influenza vaccination. *Nature* **586**, 127–132 (2020).
47. Liu, C. et al. High-resolution *HLA* typing by long reads from the R10.3 Oxford nanopore flow cells. *Hum. Immunol.* **82**, 288–295 (2021).
48. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
49. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2022).
50. Andreatta, M. et al. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat. Commun.* **12**, 2965 (2021).
51. Andreatta, M. et al. A CD4 $^+$ T cell reference map delineates subtype-specific adaptation during acute and chronic viral infections. *eLife* **11**, e76339 (2022).
52. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
53. Schmiedel, B. J. et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715 (2018).
54. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: an R-based toolkit for single-cell immune receptor analysis. *F1000Res* **9**, 47 (2020).
55. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
56. Alquicira-Hernandez, J. & Powell, J. E. Nebulosa recovers single-cell gene expression signals by kernel density estimation. *Bioinformatics* **37**, 2485–2487 (2021).
57. Borcherding, N. et al. Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. *Commun. Biol.* **4**, 122 (2021).
58. Andreatta, M. & Carmona, S. J. UCell: robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
59. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
60. Hsieh, T. C., Ma, K. H. & Chao, A. INEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).

61. Kidera, A., Konishi, Y., Oka, M., Ooi, T. & Scheraga, H. A. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J. Protein Chem.* **4**, 23–55 (1985).
62. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
63. Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E. & Friedman, N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* **33**, 2924–2929 (2017).
64. Shugay, M. et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* **46**, D419–D427 (2018).
65. Vita, R. et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
66. Zhang, W. et al. PIRD: pan immune repertoire database. *Bioinformatics* **36**, 897–903 (2020).
67. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
68. Voigt, A. P. et al. Spectacle: an interactive resource for ocular single-cell RNA sequencing data analysis. *Exp. Eye Res.* **200**, 108204 (2020).
69. Borcherding, N. & Mudd, P. Antigen-specific CD4⁺ T cells exhibit distinct transcriptional phenotypes in the lymph node and blood following vaccination in humans. Zenodo <https://doi.org/10.5281/zenodo.11395445> (2024).

Acknowledgements

We thank the research participants for their dedication to the study protocol and A. Haile and other staff members of the Washington University Infectious Disease Clinical Research Unit for their assistance with participant enrollment and follow-up. We also thank members of the Washington University Emergency Care and Research Core for their enrollment of COVID-19-infected participants. In addition, we thank A. P. Voigt of Northwestern University for providing support in the development of the interactive single-cell data portal CellPilot. This study used samples obtained from the Washington University School of Medicine's COVID-19 biorepository, which is supported by the Barnes-Jewish Hospital Foundation, the Siteman Cancer Center grant P30 CA091842 from the National Cancer Institute of the National Institutes of Health (NIH) and the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences of the NIH. The laboratory of P.A.M. was supported by US NIH grant R01AI173203 and by a grant from the Barnes-Jewish Hospital Foundation. The laboratory of A.H.E. was supported by NIH grants U01AI141990 and U01AI150747, NIAID Centers of Excellence for Influenza Research and Surveillance contracts HHSN272201400006C and HHSN272201400008C and

NIAID Collaborative Influenza Vaccine Innovation Centers contract 75N93019C00051. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIAID or NIH.

Author contributions

Conceptualization: N.B., J.S.T., A.H.E. and P.A.M. Methodology: N.B., W.K., J.Q.Z., S.A.S., P.G.T., J.S.T., A.H.E. and P.A.M. Investigation: W.K., M.Q., F.H., J.Q.Z., A. J. Sturtz, A. J. Schmitz, T.L., M.K.K., T.S., W.D.M., C.W.G., C.L., J.C.C. and S.A.T. Formal analysis: N.B. and P.A.M. Visualization: N.B. and P.A.M. Resources: C.W.G., R.M.P., J.A.O., A.H.E. and P.A.M. Data curation: N.B., W.K., J.Q.Z. and C.W.G. Funding acquisition: A.H.E. and P.A.M. Supervision: P.G.T., S.A.T., R.M.P., J.A.O., J.S.T., A.H.E. and P.A.M. Writing of the original draft: N.B. and P.A.M. Writing, reviewing and editing: all coauthors.

Competing interests

The laboratory of A.H.E. received funding under sponsored research agreements unrelated to the data presented in the current study from Emergent BioSolutions, Moderna and AbbVie. A.H.E. has received consulting and speaking fees from InBios International, Inc., Fimbrion Therapeutics, RGAX, Mubadala Investment Company, Moderna, Pfizer, GSK, Danaher, Third Rock Ventures, Goldman Sachs and Morgan Stanley. A.H.E. is the founder of ImmuneBio Consulting, LLC. N.B. is the head of computational biology at Omniscope, Inc., and has consulted for Santa Ana Bio, LLC. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41590-024-01888-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41590-024-01888-9>.

Correspondence and requests for materials should be addressed to Ali H. Ellebedy or Philip A. Mudd.

Peer review information *Nature Immunology* thanks Dirk Baumjohann, Michelle Linterman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Ioana Staicu, in collaboration with the *Nature Immunology* team.

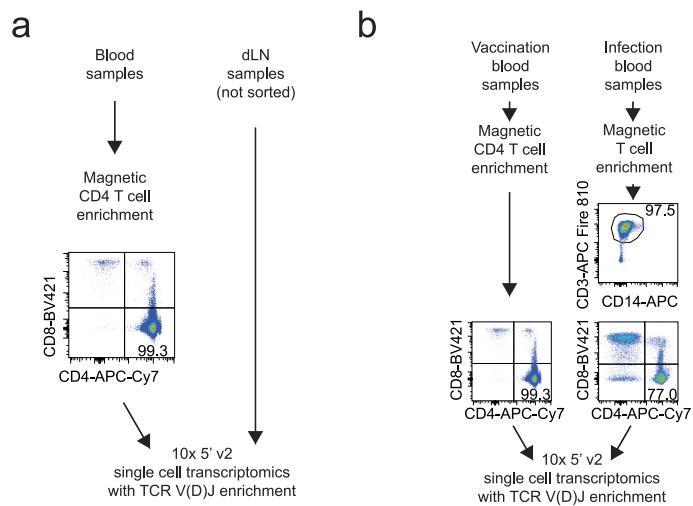
Reprints and permissions information is available at www.nature.com/reprints.

Extended Data Table 1 | Demographics of clinical cohorts

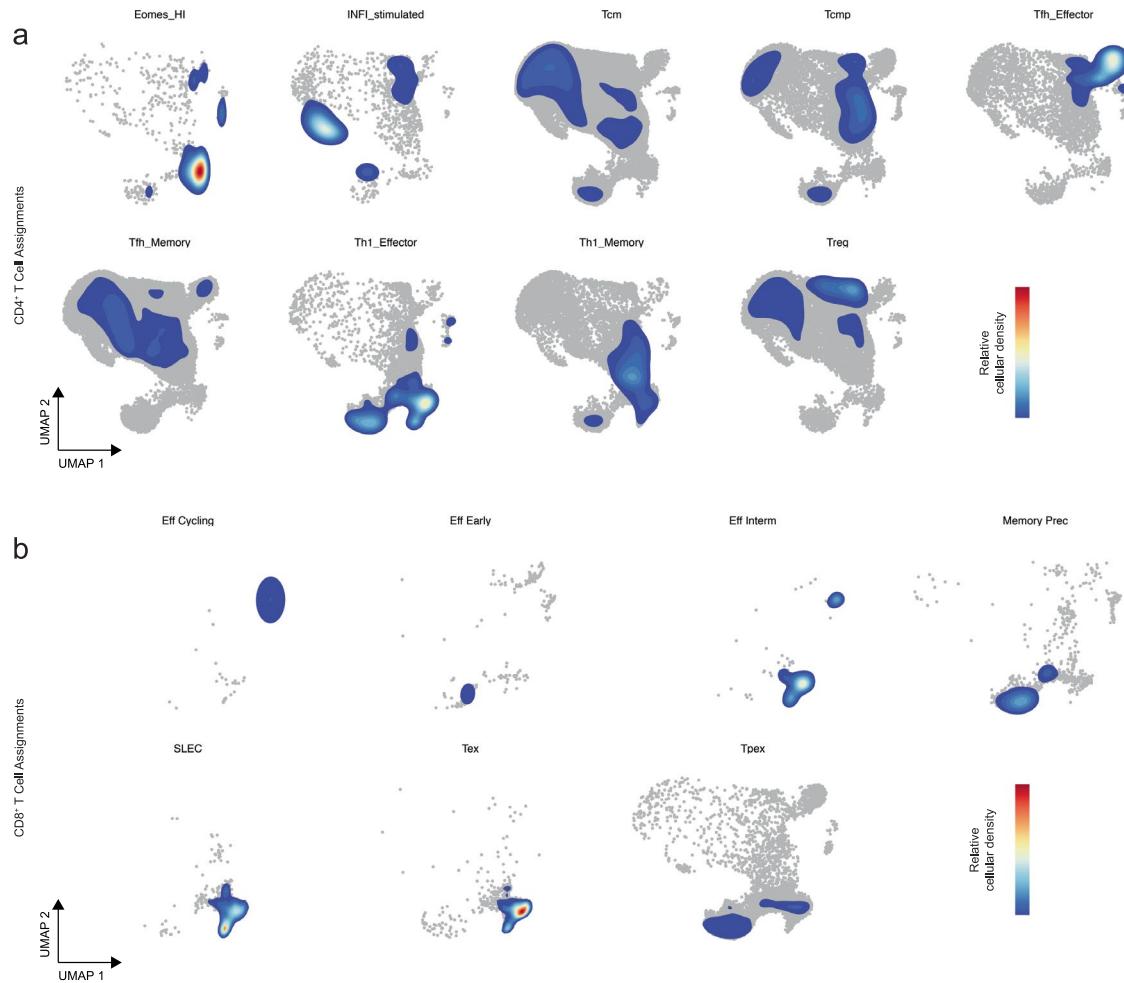
Participant	Age	Sex	Exposure	Severity	WHO severity score (0-8)	Tissue(s)
368-01a	34	male	BNT162b2			blood and lymph node
368-04	38	female	BNT162b2			blood and lymph node
368-13	34	male	BNT162b2			blood and lymph node
368-16	37	male	BNT162b2			lymph node only
368-20	48	female	BNT162b2			lymph node only
368-22	36	male	BNT162b2			blood and lymph node
350-065	73	male	SARS-CoV-2	ventilated/alive	6	blood only
350-084	67	male	SARS-CoV-2	ventilated/alive	7	blood only
350-397	53	male	SARS-CoV-2	ventilated/died	8	blood only
350-041	75	male	SARS-CoV-2	hospitalized	4	blood only
350-117	53	male	SARS-CoV-2	hospitalized	3	blood only
350-400	65	male	SARS-CoV-2	hospitalized	4	blood only

Extended Data Table 2 | HLA typing results of human donors

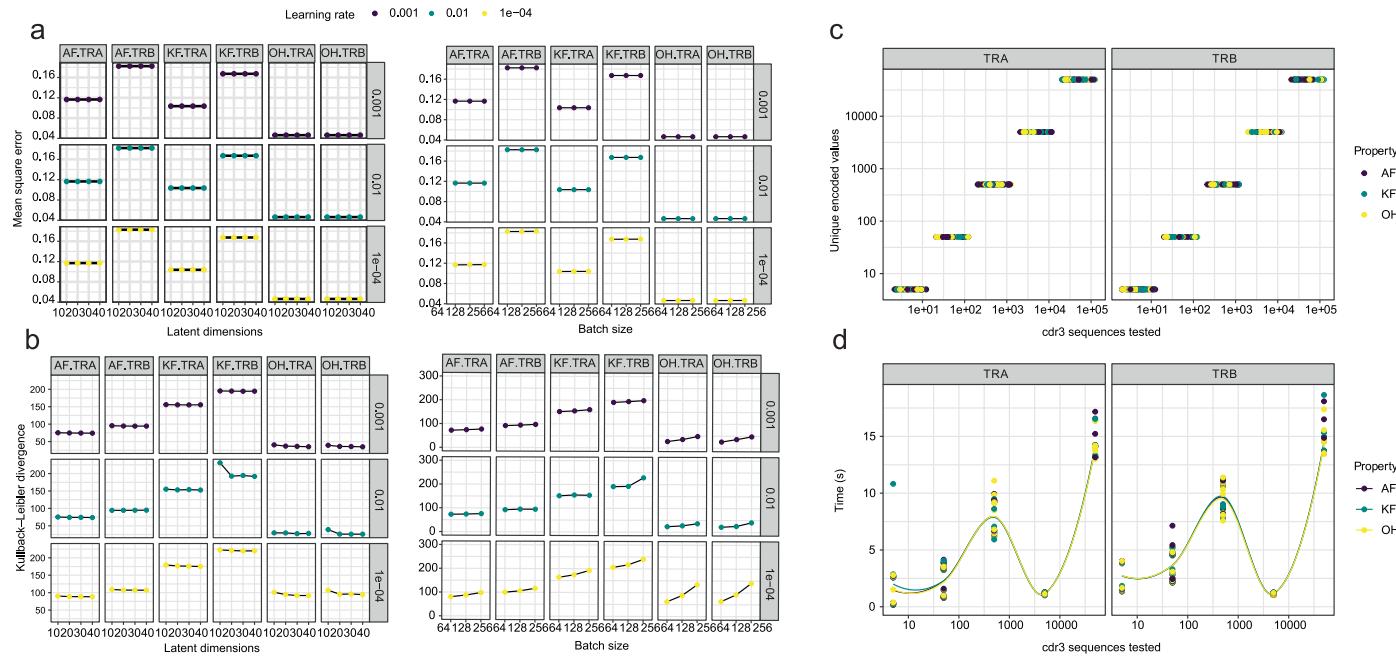
Participant	HLA-A	HLA-B	HLA-C	HLA-DPA1	HLA-DPB1	HLA-DQA1	HLA-DQB1	HLA-DRB1	HLA-DRB3,4,5
368-01a	02:01/68:01	39:01/44:03	04:01/07:02	01:03/01:03	02:01/04:01	01:02/02:01	02:02/06:02	07:01/15:01	B4*01:01/B5*01:01
368-04	01:01/03:01	08:01/15:01	03:03/07:01	01:03/01:03	04:01/04:01	03:03/05:01	02:01/03:01	03:01/04:08	B3*01:01/B4*01:03
368-13	01:01/29:02	13:02/14:02	06:02/08:02	01:03/01:03	04:01/15:01	02:01/02:01	02:02/02:02	07:01/07:01	B4*01:01/B4*01:03
368-16	02:01/03:01	35:03/51:01	04:01/15:02	01:03/01:03	02:01/04:01	01:02/01:04	05:02/05:03	14:54/16:01	B3*02:02/B5*02:02
368-20	02:01/11:01	44:02/51:01	07:04/15:02	01:03/02:01	04:01/17:01	01:01/04:01	04:02/05:01	01:01/08:01	
368-22	02:01/11:248	40:01/40:01	03:04/03:04	01:03/01:03	04:01/04:02	01:01/03:01	03:02/05:01	01:01/04:04	B4*01:03
350-065	68:01/68:01	51:01/58:02	03:04/06:02	01:03/01:03	04:01/18:01	01:05/05:05	03:01/05:01	12:01/16:02	B3*02:02/B5*02:02
350-084	02:01/02:01	27:05/45:01	02:02/16:01	01:03/03:01	04:01/105:01	03:01/05:01	02:01/03:02	03:01/04:04	B3*02:02/B4*01:03
350-397	01:01/11:01	08:01/38:01	07:01/12:03	01:03/01:03	03:01/04:01	01:02/05:01	02:01/05:02	03:01/16:01	B3*01:01/B5*02:02
350-041	02:01/03:01	07:02/56:01	01:02/07:02	01:03/01:03	03:01/04:01	03:01/05:05	03:01/03:02	04:01/12:01	B3*02:02/B4*01:03
350-117	02:01/03:01	07:02/56:01	01:02/07:02	01:03/01:03	03:01/04:01	03:01/05:05	03:01/03:02	04:01/12:01	B3*02:02/B4*01:03
350-400	02:01/03:01	15:01/44:02	03:03/05:01	01:03/01:03	04:01/04:02	03:01/03:01	03:02/03:02	04:01/04:04	B4*01:03



Extended Data Fig. 1 | Sample preparation diagram and representative flow cytometry plots of cell purity following magnetic cell enrichment. a. Blood and dLN samples from BNT162b2 mRNA vaccinated cohort. **b.** Peripheral blood mononuclear cell enrichment strategy for BNT162b2 mRNA vaccinated or SARS-CoV-2 infected donors.

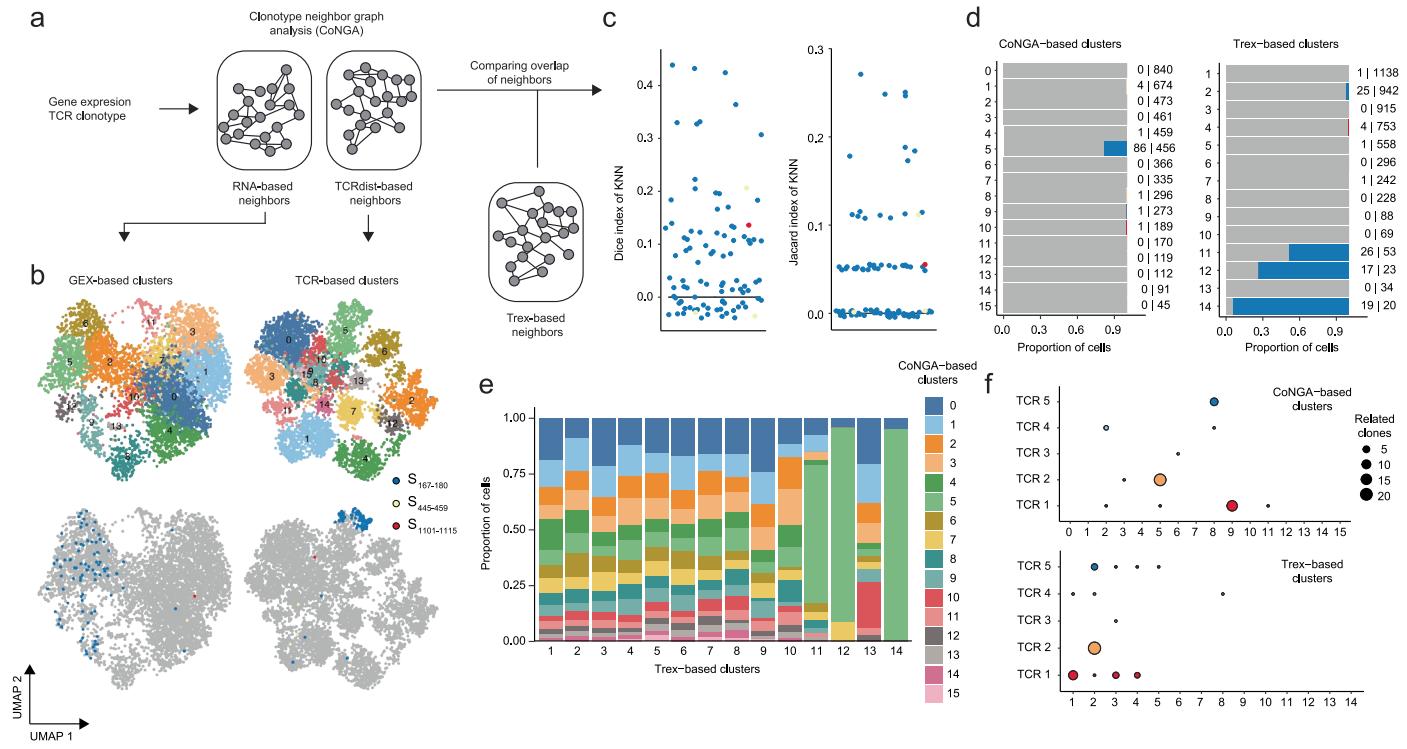


Extended Data Fig. 2 | Reference-based T cell annotations for UMAP in Fig. 1b. **a.** Density plots showing the relative distribution of ProjecTIL-based CD4⁺ T cell labels. **b.** Density plots showing the relative distribution of ProjecTIL-based CD8⁺ T cell labels. Individual gray dots indicate individual cells matching the label.



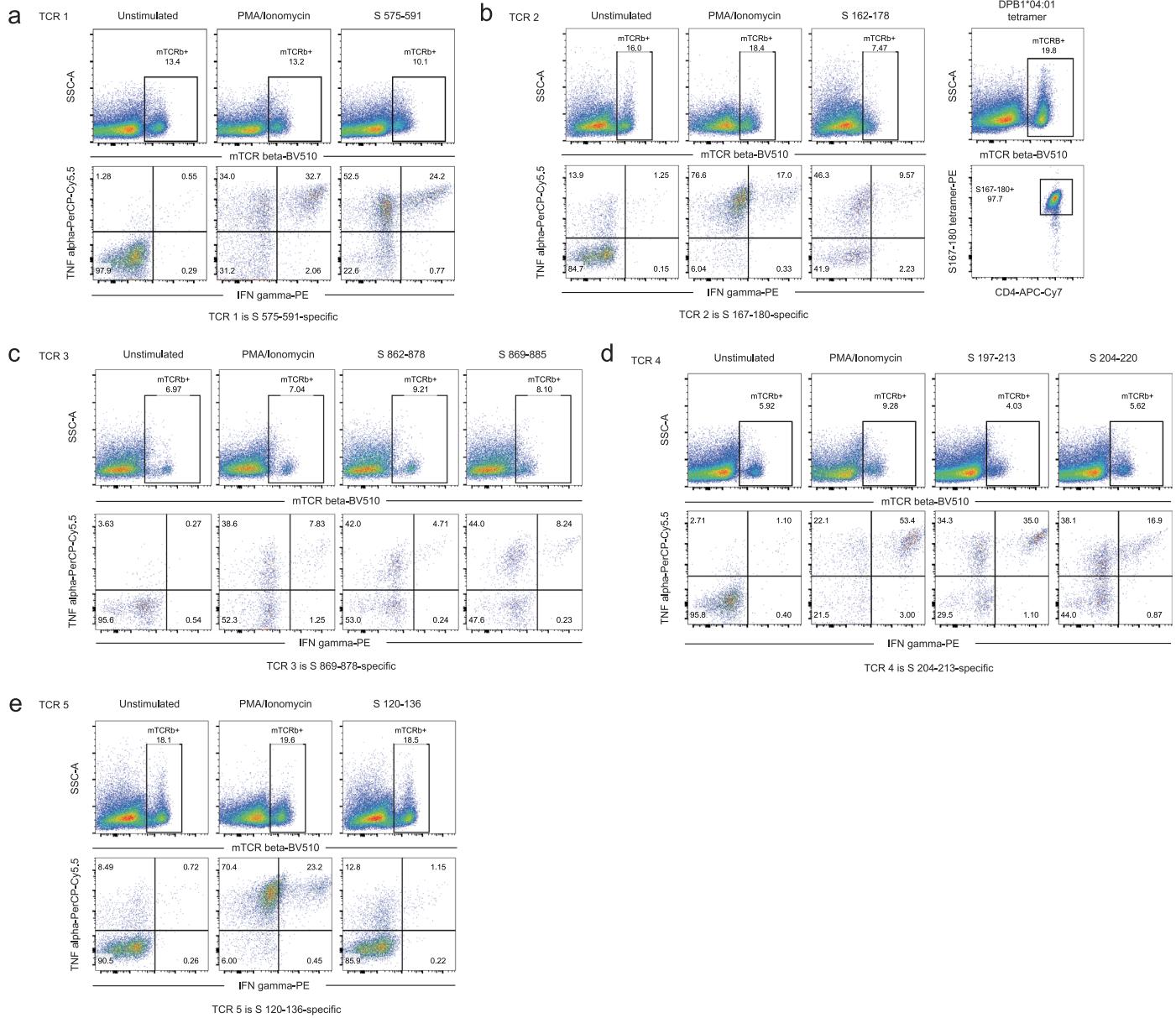
Extended Data Fig. 3 | Performance metrics for Trex autoencoder models by approach and chain. For the given hyperparameter, models were trained on 2e5 random sequences with 10 epochs for minimal Kullback-Leibler divergence value. **a.** Mean square error of models after training varying the latent dimensions (left panel) and batch size (right panel) with different learning rates. **b.** Kullback-Leibler divergence values of models after training varying the latent

dimensions (left panel) and batch size (right panel) with different learning rates. **c.** Evaluations of fidelity of models to return unique values using novel sequences for TRA and TRB chains across all models in Trex. Novel sequences were randomly sampled and bootstrapped a total of 10 times. **d.** Distribution of computational time for model application across the models, chains, and bootstraps.



Extended Data Fig. 4 | Comparison of Trex co-embedding approach with clonotype neighbor graph analysis (CoNGA). **a.** Schematic representation of the CoNGA pipeline that generates nearest neighbors of clones using both edit-distance-based TCR networks and gene expression (GEX) networks. **b.** Resulting UMAPs for CoNGA-based dimensional reduction using gene expression or edit-distance-based TCR with denoted locations of previously identified spike-specific clones. **c.** Nearest-neighbor overlap using the Dice (left) and Jaccard (right) index of the 10 nearest neighbors defined by CoNGA and by the co-

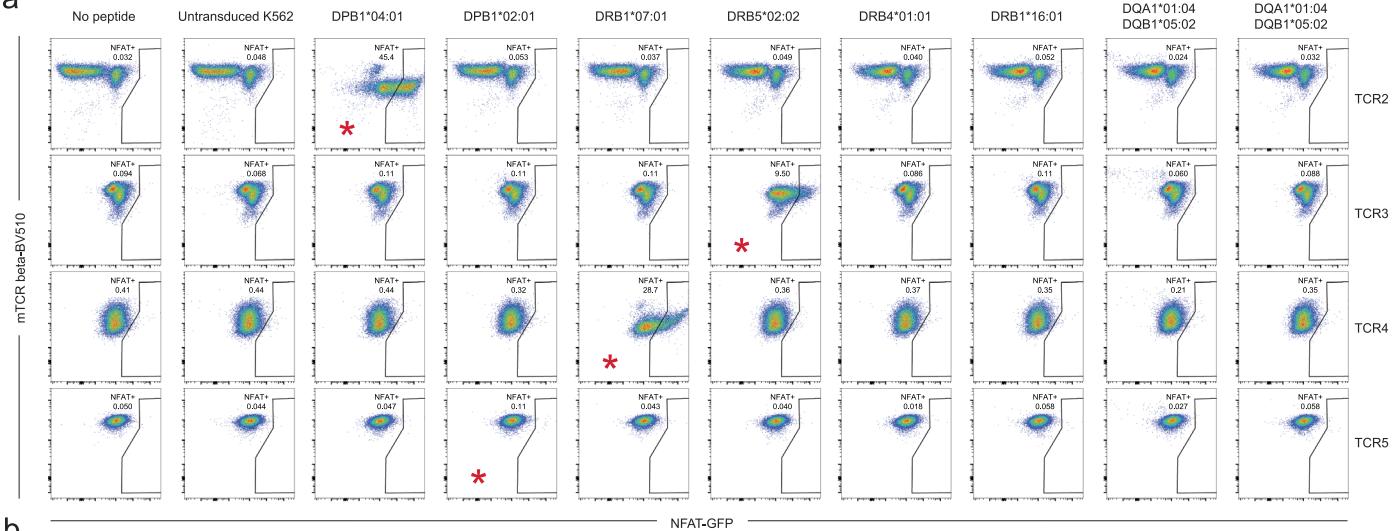
embedding with Trex. **d.** Breakdown and distribution of TCR-based clusters using CoNGA TCR output or Trex latent dimensions. Blue colored data indicate the relative proportion of clusters with spike-specific clones with a summary of the graphed values to the right of each bar chart. **e.** Trex-based latent dimensional clusters with proportion filled by the respective CoNGA TCR-based clusters. **f.** Distribution and relative size of the candidate TCRs and related sequences (edit distance ≤ 2) selected in Fig. 3 for both the CoNGA-based TCR clusters (upper panel) and Trex-based clusters (lower panel).



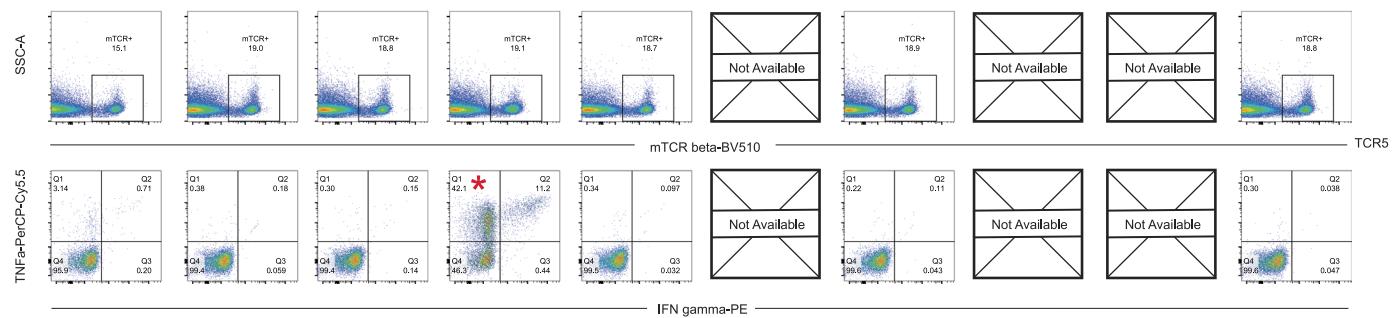
Extended Data Fig. 5 | Confirmation of TCR candidates' specificity for SARS-CoV-2 spike. Each TCR candidate's variable gene regions were cloned with murine T cell receptor (mTCR) constant regions into a retroviral transduction vector and resultant retroviruses were used to transduce primary human CD4⁺ T cells. Positive results from intracellular cytokine stain mapping of the spike protein with overlapping peptides are shown. Gating was first performed on total live single cells, then on CD3⁺CD4⁺ T cells, and finally on mTCR beta chain (mTCRb) positive candidate TCR-transduced cells. Unstimulated background

cytokine expression, positive control phorbol 12-myristate 13-acetate (PMA) and Ionomycin cytokine expression, and top cytokine expression to individual 17-mer peptides used for total spike proteome mapping are shown for each TCR candidate (a-e). Representative surface stain of unstimulated TCR2-transduced CD4⁺ T cells with the S₁₆₇₋₁₈₀ DPB1*04:01 HLA-class II tetramer is shown (right panel in b). Each experiment shown is representative of two independent TCR transduction and mapping experiments.

a

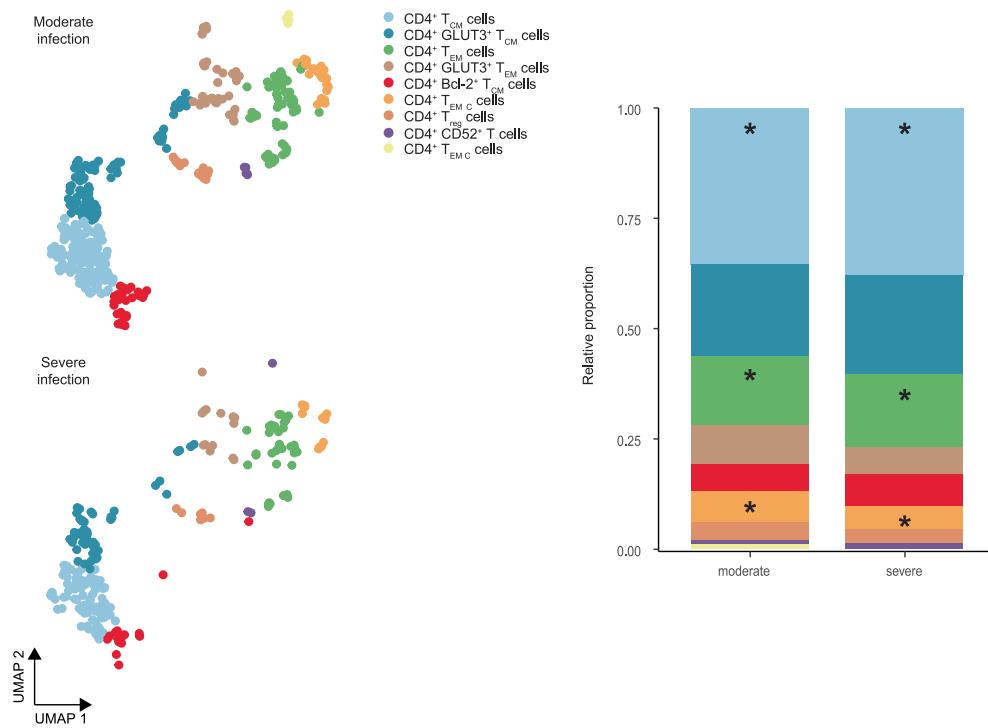


b



Extended Data Fig. 6 | Confirmation of TCR candidate HLA restriction. NFAT-GFP reporter Jurkat T cells transduced with candidate TCR expressing retrovirus were sort purified and maintained as clonal cell lines. **a.** Reporter Jurkat lines or **b.** Transduced primary human CD4⁺ T cells were co-cultured with spike peptides identified in Extended Data Fig. 5 presented in the context of various K562-based aAPC cell lines expressing single HLA class II alleles. Cells were gated on total

live single cells, then on CD3⁺ cells. In **b.**, the top panels show the frequency of retrovirally transduced (murine TCR beta constant region expressing, mTCR⁺) primary human CD4⁺ T cells that were gated on prior to evaluation of intracellular cytokine staining in the bottom panels. Red asterisks denote positive responses for each TCR line.



Extended Data Fig. 7 | Circulating blood spike-specific CD4⁺ T cells induced early after primary SARS-CoV-2 infection were similar regardless of illness severity. Comparison of circulating blood spike-specific CD4⁺ T cells during acute (day 18 to 36 post-onset of disease symptoms) infection between donors

with moderate (350-041, 350-117 and 350-400, n = 3) versus severe (350-065, 350-084 and 350-397, n = 3) infection. Statistical significance was based on bootstrapping 1,000 times to form a null distribution. * adjusted two-tailed permutation test p-value < 0.05.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Flow cytometry data were acquired using SpectroFlo software v.2.2

Data analysis Flow cytometry data were analyzed using FlowJo v.10. HLA-typing data were analyzed using Athlon2 or TypeStream Visual v.3.0. Sequencing data were analyzed using Cell Ranger v6.0, R v4.1, Seurat v4.1.0, Trex v0.99.7, mumosa v1.4.0, PhateR v1.0.7, bluster 1.6.0, leidenAl v1.0.3, msa v1.28.0, ggseqlogo v0.1, Nebulosa v1.6.0, harmony v0.1.0, scRepertoire v1.7.0, Singler v1.6.1, ProjectTIL v2.0.3, scDblFinder v1.6.0, iNEXT v3.0.0, escape v1.4.2

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All single-cell sequencing data in the manuscript including transcriptomes and T cell receptor sequencing are available through the GEO Datasets using accession

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Biological sex was not considered in study design given the limited sample size and the greater consideration that was given to including subjects with the HLA-DPBI *04 allele, but biological sex is reported for all included subjects in Supplemental Table 1.

Reporting on race, ethnicity, or other socially relevant groupings

Race, ethnicity and other socially relevant groupings were not considered in the study design. Race and ethnicity of the larger studies from which the selected subjects were chosen are reported in (Kim and Zhou et al. 2022. *Nature*. 604:141-145) and (Mudd et al. 2020. *Sci Advances*. 6(50):eabe3024).

Population characteristics

Vaccine recipients were primarily white race subjects in their 30's and 40's. Infected subjects were primarily black race subjects with ages between 53-75. Notably, the vaccine recipients included 33% biologically female subjects and the infection subjects were all biologically male.

Recruitment

Study participants were recruited from the St. Louis metropolitan area by the Washington University Emergency Care and Research Core and by the Infectious Disease Clinical Research Unit. Recruitment for the vaccine study included all eligible adults who were receiving the two-dose initial Pfizer-BioNTech SARS-CoV-2 mRNA vaccine (BNT162b2). All subjects provided written informed consent before participation. Infected subjects were all SARS-CoV-2 PCR testing positive individuals diagnosed at Barnes Jewish Hospital in St. Louis, MO, USA during the initial wave of infections in the late spring until early fall of 2020. Verbal informed consent was obtained from each subject or their legally-authorized representative. Self-selection and recruiting biases are unlikely to affect parameters measured and reported in the current study. Study subjects were financially compensated to provide samples.

Ethics oversight

The included studies were approved by the Washington University IRB.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size was determined by study participation and HLA status and not apriori by statistical methods, but gave sufficient statistics of the effect sizes of interest.

Data exclusions

There were no pre-defined exclusion criteria. We excluded only the analysis of spike antigen-specific dLN Tfh cells at the day 110 time-point in Figure 4 due to only 10 cells being obtained at that time point from all vaccine subjects which did not provide sufficient power to analyze transcriptional differences at that time point.

Replication

Technical replicates were performed for each of the single-cell RNA sequencing samples which showed good agreement. TCR transduction and mapping experiments to confirm TCR specificity were completed in two independent replicates with consistent mapping of the resultant TCR transductants to the indicated region of SARS-CoV-2 spike in each experiment.

Randomization

Randomization was not employed or applicable to this study.

Blinding

Blinding was not employed or applicable in this observational study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

- 1) CD3 APC Fire 810; SK7; BioLegend; cat# 344858; 2 uL
- 2) CD4 PerCP; OKT4; BioLegend; cat# 317432; 1 uL
- 3) CD8 BV421; RPA-T8; BioLegend; cat# 301036; 1.25 uL
- 4) CD16 BV570; 3G8; BioLegend; cat# 302036; 2 uL
- 5) CD14 APC; M5E2; BioLegend; cat# 301808; 2.5 uL
- 6) CD19 BV750; HIB19; BioLegend; cat# 302262; 1 uL
- 7) anti-murine TCR beta chain BV510; H57-597; BioLegend; cat# 109234; 2 uL
- 8) CD28 and CD49d costimulatory antibodies, unconjugated; BD Biosciences; cat# 347690; 1 uL
- 9) CD3 PE-Cy7; UCHT1; BioLegend; cat# 300420; 0.5 uL
- 10) CD4 APC-Cy7; SK3; BioLegend; cat# 344616; 0.5 uL
- 11) CD69 BV711; FN50; BioLegend; cat# 310944; 1 uL
- 12) anti-interferon gamma PE; B27; BioLegend; cat# 506507; 1 uL
- 13) anti-tumor necrosis factor alpha PerCP-Cy5.5; MAAb11; BioLegend; cat# 502926; 2 uL
- 14) anti-IL-2 APC; 5344.111; BD Biosciences; cat# 341116; 2 uL

Validation

All commercial antibodies were validated by their manufacturers as detailed in their product information page on each manufacturer's website and titrated in the prior to use in the indicated assays by serial dilution.

Eukaryotic cell lines

Policy information about [cell lines](#) and [Sex and Gender in Research](#)

Cell line source(s)

293Vec-RD114 retroviral packaging cell line obtained from BioVec Pharma, Quebec, Canada
Human Jurkat clone E6-1 T cell line obtained from ATCC

Authentication

Cells obtained directly from company or ATCC that makes or curates the indicated cell lines.

Mycoplasma contamination

Cell lines were not tested for mycoplasma contamination. Growth rates were consistent with manufacturer's data.

Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Frozen PBMC after magnetic enrichment for CD4+ cells (Figure 1) or CD3+ cells (Figure 7) were prepared as described in the methods. TCR transductants were prepared as described in methods. These cells were magnetically-enriched CD4+ frozen PBMC that were activated in culture for 2 days with T cell activating beads (Miltenyi Biotec) prior to ficoll purification and retroviral transduction as described in the detailed methods.

Instrument

Cytek Aurora

Software

Cytek SpectroFlo was used for acquisition, FlowJo v10 was used for analysis.

Cell population abundance

Purity of post-magnetic enriched populations were determined by flow cytometry and reported in the manuscript.

Gating strategy

Gating strategies are shown in Extended Data Figure 1, Extended Data Figure 5 and Extended Data Figure 6 and in each relevant panel including flow cytometry data. Initial gating was performed using forward and side scatter parameters (FSC-A/H/W, SSC-A/H/W), and Zombie NIR dye (BioLegend) to select live singlet lymphocytes. In Extended Data Figure 1 Extended Data Figure 5 and Extended Data Figure 6b, cells were then gated on CD3+ events before the presented analyses.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.